جامعة مؤتة
عمادة الدراسات العليا

تطبيق النموذج اللوغاريتمي ذي الثلاث معالم في تقدير قدرة الفرد ومعالم الفقرات لاختبار الاختيار من متعدد

بسمة محمود حسن التعمري

رسالة
مقدمة الى
عمادة الدراسات العليا
استكمالا لمتطلبات الحصول على درجة
الماجستير في القياس والتقويم قسم علم النفس

جامعة مؤتة، ٢٠٠٣

جامعة مؤتة	بسم الله الرحمن الرحيم

عمادة الدراسات العليا	**اجازة رسالة جامعية**

تقرر اجازة الرسالة المقدمة من الطالبة بسمة محمود التعمري والموسومة بـ :

"تطبيق النموذج اللوغاريتمي ذي الثلاث معالم في تقدير قدرة الفرد ومعالم الفقرات لاختبار الاختيار من متعدد" استكمالا لمتطلبات الحصول على درجة الماجستير في القياس والتقويم.

القسم : علم النفس.

الاسم	التوقيع	التاريخ	
د. ماهر الدرابيع		٢٠٠٣/٨/٢٨	مشرفاً
د. عبداللة الصمادي		٢٠٠٣/٨/٢٨	عضواً
د. عماد الزغول		٢٠٠٣/٨/٢٨	عضواً
د. محمد المجالي		٢٠٠٣/٨/٢٨	عضواً

عميد الدراسات العليا

د. ذياب البداينة

الشكر والتقدير

الشكر الكبير والتقدير لأستاذي ومشرفي الدكتور ماهر الدرابيع الذي كان خير مشرف وناصح لي، فهو لم يتوانى عن تقديم المساعدة الكبيرة لي مهما كانت ظروفه، فكل التقدير والعرفان لمعلمي الفاضل.

كل الشكر للأساتذة في لجنة المناقشة، الدكتور عبدالله الصمادي والدكتور عماد الزغول والدكتور محمد المجالي على تفضلهم بمناقشة الرسالة وتقديمهم الرأي وما من شأنه أن ينهض بهذا العمل العلمي.

والى كل من قدم لي أيسر مساعدة للنهوض بعملي هذا له مني كل الشكر والتقدير.

بسمة محمود التعمري

جدول المحتويات

المحتوى الصفحة

قائمة الجداول

رقم الجدول   موضوع الجدول   الصفحة

الملخص

تطبيق النموذج اللوغاريتمي ذي الثلاث معالم في تقدير قدرة الفرد ومعالم الفقرات لاختبار الاختيار من متعدد

بسمة محمود حسن التعمري

جامعة مؤتة، 2003

هدفت هذه الدراسة الى تطبيق النموذج اللوغاريتمي ذي الثلاث معالم لتقدير قدرة الفرد ومعالم الفقرات وذلك لاختبار الاختيار من متعدد بإختلاف عدد البدائل، كما هدفت هذه الدراسة الى مقارنة النتائج في ضوء النظرية الكلاسيكية في القياس.

وقد تكونت عينة الدراسة من (1200) طالب وطالبة من جميع طلبة الصف العاشر الأساسي في مديريات التربية التابعة لمديرية تربية عمان الرابعة.

وقد تم استخدام برنامج (BILOG) لتقدير قدرة الفرد ومعالم الفقرات وفق النظرية الحديثة في القياس.

وقد أشارت نتائج هذه الدراسة الى عدم وجود فروق ذات دلالة احصائية في قدرة الفرد على اختبار الاختيار من متعدد باختلاف عدد البدائل وفق النظرية الحديثة في القياس، في حين دلت النتائج الى وجود فروق في قدرة الفرد وفق النظرية الكلاسيكية في القياس، لصالح الاختبار ذي الخمس بدائل.

كما أشارت النتائج الى وجود فروق ذات دلالة احصائية في صعوبة الفقرة بإختلاف عدد البدائل وفق النظرية الحديثة في القياس لصالح الاختبار ذي الثلاث بدائل، في حين أن النتائج أشارت الى عدم وجود

فروق ذات دلالة احصائية في صعوبة الفقرة بإختلاف عدد البدائل وفق النظرية الكلاسيكية في القياس.

وأشارت النتائج الى وجود فروق ذات دلالة احصائية في تمييز الفقرة على اختبار الاختيار من متعدد بإختلاف عدد البدائل وفق النظرية الحديثة في القياس وذلك لصالح الاختبار ذي الخمس بدائل، والى وجود فروق ذات دلالة احصائية في تمييز الفقرة وفق النظرية الكلاسيكية في القياس لصالح الاختبار ذي الأربع بدائل.

وأخيرا أشارت النتائج الى وجود فروق ذات دلالة احصائية في تخمين الفقرة على اختبار الاختيار من متعدد بإختلاف عدد البدائل في النظرية الحديثة في القياس وذلك لصالح الاختبار ذي الثلاث بدائل.

Abstract

The application of the Three-Parameter Logistic Model in Estimating Person ability and Item parameters for the multiple choice test

Bassmah Mahmoud Altamary

Mu'tah University, 2003

The aim of this study is to estimate the person ability and item parameters for the multiple choice test with different options regarding to the Classical Test theory and the Item Response theory when apply the Three-Parameter Logistic Model.

The sample of the study consisted (1200) tenth grade students from different schools in the fourth Amman Department of Education.

BILOG program was used to estimate person ability and Item-parameters regarding the "IRT" module.

The results indicated no statistical significant differences in person ability with different alternatives in the multiple-choice exam regarding the "IRT" module, but it was indicated a significant differences regarding the "CTT" in favor of five options.

Also, the results indicated statistical significant differences in item difficulty regarding to "IRT" in favor of three options, but it was found no statistical significant differences regarding to "CTT", and the results indicated statistical significant differences in item discrimination regarding to "IRT" in favor to five option items and in favor of four options regarding "CTT". Finally, the results indicated statistical significant differences in item guessing regarding "IRT" in favor of three options.

الفصل الأول

خلفية الدراسة وأهميتها

مقدمة

لقد شهد ميدان القياس والتقويم تطورا كبيرا في مجال تطوير الاختبارات خلال العقود الأخيرة (Baker, 1985)، وتعتبر النظرية الحديثة في القياس أو ما تسمى بنظرية إستجابة الفقرة أكبر نقطة تحول في هذا التطور(Warm, 1978) بسبب ما قدمته من طرق ذات فعالية كبيرة فيما يتعلق ببناء وتصحيح وتحليل الاختبارات مقارنة بما قدمته النظرية التقليدية في القياس.

هناك ثلاثة نماذج تعتبر الأكثر استعمالا في مجال البحث ضمن نظرية استجابة الفقرة، أولها النموذج اللوغاريتمي ذو المعلمة الواحدة ويسمى بنموذج راش حيث يعتبر هذا النموذج الأكثر استعمالا من بين نماذج النظرية الأخرى (Harris, 1989). أما النموذج الثاني ضمن هذه النظرية فهو النموذج اللوغاريتمي ذي المعلمتين، ويعتبر هذا النموذج أقل عمومية واستعمالا من نموذج راش.

ويعتبر النموذج اللوغاريتمي ذي الثلاث معالم هو النموذج الثالث في هذه النظرية، ويتطلب هذا النموذج مقارنة بالنماذج السابقة عددا كبيرا من المفحوصين وعددا كبيرا من فقرات الاختبار.

ولذا حاولت هذه الدراسة تطبيق النموذج اللوغاريتمي ذي الثلاث معالم على اختبار إختيار من متعدد وذلك بعدد بدائل مختلف (3,4,5) على عينة الدراسة؛ وذلك في ضوء قدرة الفرد ومعالم الفقرات (تمييز الفقرة وصعوبة الفقرة وتخمين الفقرة)، كما حاولت هذه الدراسة مقارنة النتائج في ضوء النظرية الحديثة في القياس والنظرية التقليدية في القياس.

أهمية الدراسة :

تهدف هذه الدراسة الى تطبيق النموذج اللوغاريتمي ذي الثلاث معالم في تقدير قدرة الفرد ومعالم الفقرات وذلك لاختبار الاختيار من متعدد تعزى الى اختلاف عدد البدائل (3,4,5)، وتأتي أهمية هذه الدراسة من أن الدراسات السابقة استندت الى تقييم نتائجها باستخدام الطرق الإحصائية الوصفية، بحيث قدمت للقاريء بيانات وصفية فقط، لذلك بات من الضروري أن نوظف هذه البيانات في فحص فرضيات احصائية تتعلق بقدرة الفرد على الاختبار ومعرفة معالم الفقرات وهي: صعوبة الفقرة وتمييز الفقرة والتخمين. فقد أوصى هارول (Harwell, 1997) بضرورة الدراسة التحليلية لنتائج الدراسات التي تعتمد على طريقة المحاكاة في توليد الإستجابات للفقرات باستخدام التصاميم التجريبية، وليس الإكتفاء بالتحليلات الوصفية.

كما تهدف هذه الدراسة الى مقارنة النتائج عند تطبيق النموذج اللوغاريتمي ذي الثلاث معالم على اختبار الاختيار من متعدد تعزى الى اختلاف عدد البدائل في تقدير قدرة الفرد ومعالم الفقرات بالنموذج التقليدي في القياس أو ما يسمى بالنظرية الكلاسيكية في القياس.

أسئلة الدراسة :

تحاول هذه الدراسة الإجابة عن الأسئلة التالية:

السؤال الأول: هل هناك فروقا ذات دلالة إحصائية في قدرة الفرد على اختبار الاختيار من متعدد تعزى الى اختلاف عدد البدائل وفق النظرية التقليدية في القياس؟

السؤال الثاني: هل هناك فروقا ذات دلالة إحصائية في قدرة الفرد على اختبار الاختيار من متعدد تعزى الى اختلاف عدد البدائل وفق النظرية الحديثة في القياس؟

السؤال الثالث :هل هناك فروقا ذات دلالة احصائية في صعوبة الفقرة على اختبار الاختيار من متعدد تعزى الى اختلاف عدد البدائل وفق النظرية التقليدية في القياس؟

السؤال الرابع: هل هناك فروقا ذات دلالة احصائية في تمييز الفقرة على اختبار الاختيار من متعدد تعزى الى اختلاف عدد البدائل وفق النظرية التقليدية في القياس؟

السؤال الخامس: هل هناك فروقا ذات دلالة إحصائية في صعوبة الفقرة على اختبار الاختيار من متعدد تعزى الى اختلاف عدد البدائل وفق النظرية الحديثة في القياس ؟

السؤال السادس: هل هناك فروقا ذات دلالة إحصائية في تمييز الفقرة على اختبار الاختيار من متعدد تعزى الى اختلاف عدد البدائل وفق النظرية الحديثة في القياس ؟

السؤال السابع: هل هناك فروقا ذات دلالة إحصائية في تخمين الفقرة على اختبار الاختيار من متعدد تعزى الى اختلاف عدد البدائل وفق النظرية الحديثة في القياس؟

ويمكن صياغة سؤال عام للدراسة على النحو التالي:

هل تختلف تقدير قدرة الفرد ومعالم الفقرات باختلاف عدد البدائل في اختبار الاختيار من متعدد في ضوء النظرية التقليدية في القياس والنظرية الحديثة في القياس عند تطبيق النموذج اللوغاريتمي ذي الثلاث معالم ؟

الفصل الثاني

الإطار النظري و الدراسات السابقة

تفترض الاختبارات النفسية والتربوية وجود سمات أو خصائص معينة يشترك فيها جميع الأفراد، ولكنهم يختلفون في مقدارها، ويمكن الاستدلال على مقدار هذه السمات من السلوك الملاحظ للفرد الذي يتمثل في استجابته لفقرات الاختبار، ولذلك تسمى هذه السمات بالسمات الكامنة والأمر الذي يهمنا في القياس النفسي التربوي هو تحديد مقدار السمات الكامنة وراء أداء الأفراد في الاختبارات المختلفة للإفادة من ذلك في التنبؤ بسلوكهم في مواقف مماثلة ومن أجل إتخاذ القرارات المناسبة بشأنهم (علام، 1986) .

لذا فأدوات القياس لمثل هذه السمات تكون على شكل اختبارات ومقاييس، كاختبارات القابلية و التحصيل ومقاييس الاتجاهات والقدرات.

من هنا كان دأب علماء القياس يتركز على تطوير أدوات القياس للحصول على نتائج تتمتع بالدقة والتمثيل الكافي لمجال السلوك المراد قياسه؛ أي بمعنى أخر أن تتحقق صفة الموضوعية في القياس بحيث لا يتأثر قياس أي ظاهرة سلوكية باختلاف الأداة المستخدمة والأفراد الذين تم قياسهم بهذه الأداة. ومن خلال مراجعة الأدب السابق فإن هناك تعريفات مختلفة للقياس فقد عرفه كامبل (Campbell & Fiske, 1959) على أنه الإشارة العددية للخصائص المتمثلة في الأنظمة المادية ضمن قواعد معينة تحكم هذه الخصائص. في حين عرفه ستيفنز (Stevens, 1951) بأنه إعطاء قيم رقمية لخاصية ما بحيث يمثل الرقم مقدار امتلاك الفرد أو الشيء لتلك الخاصية وفق قواعد معينة.

وهكذا فأنه يمكن القول بأن مصطلح القياس يشير الى مجموعة الإجراءات التي تتضمن تحديد وتعريف ما يجب قياسه وترجمته إلى معلومات يسهل وصفها بمستوى مقبول من الدقة؛ أي أن القياس يصف السلوك وصفا كميا بناء على أن القياس عدد ووحدة (عودة، 1993).

وهذا يعني أن المعلومات أو البيانات المتعلقة بسلوك معين يمكن جمعها بأدوات توصف بأنها أدوات قياس إذا وفرت معلومات كمية بطريقة مباشرة كالاختبارات أو بطريقة غير مباشرة كالاستبيانات التي تتكون من مجموعة من الفقرات التي تحمل في مضمونها أهدافا تخدم أغراضها.

وفي العموم فإن هناك إسلوبان لاختيار الفقرات أحدهما يعتمد على النظرية التقليدية (Classical Test Theory, CTT)، والآخر يعتمد على النظريةالحديثة في القياس (Modern Test Theory, IRT)، فالنظرية التقليدية تفرز فقرات تتمتع بخصائص معينة اعتمادا على معامل الصعوبة والتمييز وكذلك فإن النظرية الحديثة تفرز أيضا فقرات تتمتع بخصائص معينة اعتمادا على معامل الصعوبة والتمييز وكذلك معامل التخمين، وقد تتفق النظريتان في قبولها أو رفضها لهذه الفقرات وفقا لخصائصها و معاييرها.

وفيما يلي تعريف بهاتين النظريتين ليتضح دورهما في إختيار الفقرات في الاختبار الواحد وفق خصائص معينة تحددها كل نظرية.

أولاً: النظرية التقليدية: (Classical Test Theory)

النظرية التقليدية في القياس (CTT) هي نموذج بسيط يحتوي تعاريف عن العلامات الحقيقية والخطأ العشوائي في القياس والعديد من الافتراضات التي تهتم بالارتباطات بين العلامات الحقيقية والخطأ.

وتقوم النظرية التقليدية في القياس على افتراض أن الفرد يمتلك قدرا معينا من السمة يمكن قياسها باستخدام اختبار معين، وبأن العلامة التي يحصل عليها في الاختبار هي دالة (Function) لقدرته الحقيقية على تلك السمة، ومقدار الخطأ العشوائي اللاتي تؤثران على العلامة الملاحظة للفرد، أي أن العلامة الملاحظة تتكون من جزأين، هما: العلامة الحقيقية (T) وهو مقدار يفترض ثباته، والجزء الآخر هو مقدار الخطأ العشوائي حيث يتوزع هذا الخطأ توزيعا عشوائيا بوسط حسابي مقداره صفر وانحراف معياري لا يساوي الواحد الصحيح. والعلاقة بين العلامة الحقيقية ومقدار الخطأ العشوائي والعلامة الملاحظة تمثل بالعلاقة التالية:

$$X=T+E \qquad \text{(Lord\&Novick,1968)}$$

حيث X العلامة الملاحظة.

T العلامة الحقيقية.

E مقدار الخطأ العشوائي.

ويقصد بمقدار الخطأ العشوائي هو الفرق بين العلامة الحقيقية والعلامة الملاحظة، والأخطاء هي عوامل تجعل العلامة الملاحظة أحيانا أعلى وأحيانا أقل من العلامة الحقيقية وبشكل عشوائي، ويسمى الخطأ بمثل هذه الحالة بالخطأ العشوائي (Random Error)، وقد يكون التأثير ثابتا أي في إتجاه واحد إما زيادة أو نقصان، ويسمى في هذه الحالة بالخطأ المنتظم (Systematic Error).

ويؤثر الخطأ العشوائي في العلامة الملاحظة بطريقة عشوائية أي ليس بنفس الإتجاه كما في الخطأ المنتظم وبالتالي ستكون نتيجة القياس غير صادقة وغير دقيقة، لذا عند الإشارة لخطأ القياس في هذه النظرية يقصد به مقدار الخطأ العشوائي.

ولقد بين جاليكسان (Gulliksen) المشار إليه في (Crocker and Algina , 1986) مباديء وأسس هذه النظرية، والتي تتضمن مجموعة من الافتراضات أهمها :

1- تمثل العلامة الظاهرية (X) حاصل جمع العلامة الحقيقية (T) ومقدار الخطأ في القياس (E) علما بأن مقدار الخطأ يمكن أن يكون بالزيادة أو النقصان، والثابت في المعادلة هو العلامة الحقيقية.
2- لا يمكن معرفة العلامة الحقيقية أو قياسها بل نستدل عليها من خلال العلامة الظاهرية.
3- لا يوجد ارتباط بين العلامات الحقيقية وأخطاء القياس لعدد من المفحوصين على نفس الاختبار، أي أنه ليس من الضروري أن يكون مقدار أخطاء القياس للمفحوصين ذوي العلامات الحقيقية المرتفعة أكبر من تلك لذوي العلامات الحقيقية المتدنية أو عكس ذلك.
4- لا يوجد ارتباط بين أخطاء القياس في أي اختبارين مختلفين لنفس المفحوص (ويمكن لهذا الإفتراض أن لا يتحقق إذا كان الأداء على الاختبار يتأثر كثيرا بظروف التطبيق) .
5- إن خطأ القياس هو خطأ عشوائي، أما الخطأ المنتظم فهو متعلق بصدق القياس أو صدق الاختبار، إلا أن الخطأ العشوائي هو الذي يحدد دقة القياس أو ما يسمى بثبات الاختبار.

6- تفترض هذه النظرية أن تباين أخطاء القياس هو نفسه لجميع الأفراد الذين يطبق عليهم الاختبار، لأن مقدار الخطأ العشوائي هو دالة لمعامل الثبات.

فالعلامة الحقيقية (T) وخطأ القياس (E) هما علامتان نظريتان غير ظاهرتان، فالعلامة الحقيقية هي متوسط العلامات الناتجة من إعادة تطبيق نفس الاختبار بعدد لا نهائي من المرات المستقلة على نفس المفحوص وهذا السبب هو الذي جعلها قيمة نظرية.

وهذه العلامة (T) لن تعكس القيمة الحقيقية لدرجة امتلاك الفرد للسمة المقاسة ما لم يتمتع الاختبار بالصدق التام، أي عندما يقيس الاختبار ما بني لأجله فقط (Allen and Yen, 1979).

أما الميزة الرئيسية التي يمكن أن توصف بها هذه النظرية فهي بساطتها وسهولة تطبيقها (Hambelton and Jones, 1993) لأن معظم تركيزها ينصب على خصائص سيكومترية بسيطة مثل معامل الصعوبة والتمييز.

بالرغم من المميزات السابقة والتي تتميز بها النظرية التقليدية في القياس إلا أن هناك بعض القصور يتمثل في:

1- تمتاز خصائص الفقرات التي تبنى وفق أسس هذه النظرية بكونها غير مستقلة عن خصائص المفحوصين، وكذلك خصائص المفحوصين غير مستقلة عن خصائص عينة الفقرات في ذلك الاختبار. أي أن الاختبارات المبنية وفق اسس هذه النظرية والمفاهيم السيكومترية المرتبطة بها من مثل معاملات الصعوبة والتمييز تختلف باختلاف خصائص أفراد العينة المستخدمة في حساب هذه المعاملات (Hambelton and Swaminathan, 1985).

2- عند مقارنة المفحوصين على قدرة أو سمة معينة باستخدام مجموعة من فقرات الاختبار فان هذه المقارنة تحددها الاوضاع والظروف التي تم فيها تطبيق نفس فقرات الاختبار، وعلى هذا فان المقارنة لا يمكن التعامل معها على انها اخذت من اختبارات معيارية طبقت في ظروف معيارية.

3- تقوم فكرة ثبات المقاييس في النظرية التقليدية على أساس الصور المتوازية للاختبار، التي يصعب الحصول عليها عمليا. وذلك لأن الأفراد في التطبيق الثاني للاختبار لا يمكن أن يكونوا هم أنفسهم في التطبيق الأول بسبب تأثرهم بعوامل التذكر أو النسيان أو قلق الاختبار وغيرها من العوامل الأخرى (Hambelton and Swaminathan , 1985).

4- تفترض هذه النظرية أن تباين اخطاء القياس (الذي هو دالة معامل الثبات) هو نفسه لجميع المفحوصين (Lord, 1984) وهذا غير مقبول، فأخطاء القياس للعلامات على اختبار صعب يكون أعلى للطلبة ذوي القدرة المتدنية منها لدى الطلبة ذوي القدرة المتوسطة أو العالية (Hambelton and Swaminathan, 1985).

5- لا توفر هذه النظرية أي قاعدة للتنبؤ في كيفية اداء الفرد على أي فقرة من فقرات الاختبار، أي بمعنى أننا لا نستطيع من خلال هذه النظرية الاجابة عن سؤال: ما احتمال ان يجيب الفرد اجابة صحيحة على فقرة اختبار محددة؟

لهذه الأسباب فان علماء النفس يبحثون عن نظريات ونماذج بديلة للقياسات الموضوعية والتي يتطلعون الى ان تتميز بما يلي:

1- أن لا تكون خصائص الفقرة معتمدة على المجموعة.

2- أن لا تكون العلامات التي تصف الأفراد معتمدة على الاختبار.

3- أن تكون نموذج يعبر عن مستوى الفقرة اكثر من أن يعبر عن الاختبار ككل.

4- أن تكون نموذج لايحتاج لوجود نماذج متكافئة من الاختبار لتحقيق الثبات.

5- أن تكون نموذج يزودنا بقياس واضح لكل من قدرة الفرد ومعالم الفقرات.

ثانيا: النظرية الحديثة في القياس: (Modern Test Theory)

يرى بعض العلماء مثل (Hambelton and Swaminthan, 1985) أنه بالإمكان الحصول على كل ما سبق من خصائص بإيجاد نظرية اختبار تعرف بنظرية استجابة الفقرة (Item Response Theory, IRT) كما تسمى أيضا بالنظرية الحديثة في القياس (Modern Test Theory).

في عام 1952 نشر العالم لورد (lord) رسالة الدكتوراة التي قدم فيها النظرية الحديثة في القياس كنموذج من نماذج القياس. بحيث قام بوضع أسس وافتراضات وصاغها على شكل نظرية سماها بنظرية منحنى خصائص الفقرة.

وتقوم هذه النظرية (IRT) على مسلمتين أساسيتين هما:

1- يتحدد احتمال اجابة الفرد على الفقرة بدلالة السمة الكامنة لديه أو ما يسمى بقدرة الفرد (Person Ability).

2- يمكن أن توصف العلاقة بين أداء الفرد على الفقرة وقدرته من خلال منحنى خصائص الفقرة (Item Characteristic Curve, ICC) وهذه العلاقة تتحدد بزيادة السمة لدى الفرد مما يزيد من احتمالية الاجابة الصحيحة على الفقرة لديه.

تحتوي جميع نماذج هذه النظرية على واحد أو أكثر من المعالم التي تصف الفقرة كصعوبة الفقرة وتمييز الفقرة وتخمين الفقرة بالإضافة الى قدرة الفرد. والخطوة الأولى في تطبيق هذه النظرية هي تقدير هذه المعالم.

ويوفرنموذج النظرية الحديثة في القياس علاقة تربط بين أداء الفرد على الاختبار والسمات الكامنة التي يفترض بأن تؤثر في هذا الأداء، حيث تصف النظرية العلاقة باقتران رياضي لوغاريتمي؛ ولهذا كانت النماذج المنبثقة عنها نماذج رياضية لوغاريتمية.

افتراضات النظرية الحديثة في القياس:

تقوم نظرية استجابة الفقرة على افتراضين أساسيين هما:

أولا: أحادية البعد (Unidimensionality):

ويشير هذا الافتراض إلى وجود قدرة أو سمة واحدة فقط تفسر اداء الفرد على الفقرة وهي ما تسمى بقدرة الفرد، ولكن بسبب عوامل أخرى ربما تؤثر في الأداء بجانب قدرة الفرد كالعوامل الشخصية والعقلية وعوامل لها علاقة بظروف تطبيق الاختبار كمستوى الدافعية وقلق الاختبار والقدرة على الإنجاز السريع والميل الى التخمين، لذلك فهذا الافتراض قد لايتحقق دائما بشكل مؤكد. اذن ماالذي نحتاجه لتحقيق هذا الإفتراض؟

يرى هامبلتون (Hambleton, 1989) انه بالرغم من وجود مثل هذه العوامل فانه يتطلب وجود عامل اساسي يؤثر في الاداء على الاختبار وهو ما يشار إليه بالقدرة التي تقيسها الفقرة، ومع ذلك فإنه ليس من الضرورة أن تكون موجودة بالأصل، ولكن من المتوقع ان تتغير علامات القدرة بمرور الوقت بسبب التعلم و النسيان وعوامل أخرى.

في حين أن النماذج التي تفترض وجود أكثر من قدرة لتوضيح أداء الفرد تسمى بالنماذج متعددة الأبعاد (Multi Dimensional)، وهذه النماذج معقدة كثيرا لم يتم تطويرها بشكل جيد (Mcdonalid, 1974).

ثانيا: الاستقلال المحلي (Local independence):

ويقصد بهذا الافتراض أن أداء الأفراد على فقرات الاختبار يكون مستقلا إحصائيا عندما يؤخذ مستوى القدرة بعين الإعتبار، وهذا يتضمن أن أداء الفرد على فقرة من فقرات الاختبار لا يؤثر على أدائه على أية فقرة أخرى على افتراض أن كل فقرة مستقلة في بنائها ولا توحي بالاجابة عن الفقرة التي تليها، بالإضافة الى أن كل فقرة تحتاج مستوى معين من القدرة.

وهذا يقودونا الى اعتبار أن القدرة التي يملكها الفرد هي العامل الوحيد المؤثر في استجابته على الفقرة، ومجموعة القدرات هذه تمثل المجال الكامن الكامل (The Complete Latent Space) وعندما يتحقق افتراض أحادية البعد فإن الفضاء الكامن الكامل يحتوي على قدرة واحدة، ومن الملاحظ أن هذين الافتراضين متلازمان (أحادية البعد والاستقلال المحلي)، بمعنى أنه إذا تحقق الافتراض الاول فان الافتراض الثاني يتحقق أيضا (Hambleton, 1989)، وبالمنطق سيكون هذين المفهومين متكافئان(Lord1980,lord&Novick, 1968).

ولكن يمكن تحقيق الاستقلال المحلي حتى لو لم تكن مجموعة البيانات أحادية البعد لأن الاستقلال المحلي يمكن تحقيقه كلما كان الفضاء الكامن الكامل محددا؛ اي بمعنى عندما تكون جميع ابعاد القدرة التي تؤثر على الافراد قد اخذت بعين الاعتبار.

وعلى العكس من ذلك فإنه ليس بالإمكان تحقيق الاستقلال المحلي عندما يكون الفضاء الكامن الكامل غير محدد؛ أي أنه ليس من الممكن تحقيق الاستقلال المحلي عندما تحتوي فقرة الاختبار على إشارة للإجابة الصحيحة، أو عندما تزودنا الفقرة بايحاءات بحيث تساعدنا على الإجابة عن فقرة أخرى، وفي هذه الحالة من الممكن أن يكتشف بعض المفحوصين هذه الإشارة والبعض الآخر لا يمكن أن يكتشفها وهنا تصبح القدرة على كشف الإشارة هي البعد الذي يفحص وليست القدرة.

وبناءا على هذين الافتراضين فقد اشتملت نظرية استجابة الفقرة على عدد من النماذج الرياضية بحيث تختلف باختلاف عدد المعاملات التي يأخذها النموذج بعين الاعتبار، في حين أن كل هذه النماذج تهدف إلى تحديد العلاقة بين أداء الفرد على الاختبار وبين القدرات التي تكمن وراء هذا الأداء (Hambelton and Jones, 1993).

ولعل التمييز الأساسي بين نماذج استجابة الفقرة أحادية البعد الشائعة هو في عدد المعالم المستخدمة لوصف الفقرات.

ويرجع اختيار النموذج الى المستخدم، ولكن هذا الاختيار يتضمن افتراضات حول البيانات التي من الممكن أن تتحقق بفحص كيفية تفسير النموذج لنتائج الاختبار الملاحظة.

نماذج النظرية الحديثة في القياس:

هناك ثلاثة نماذج شائعة في نظرية استجابة الفقرة والتي تعتبر الأكثر استعمالا في مجال البحث ضمن نظرية إستجابة الفقرة أحادية البعد، وهي:

أولا: النموذج أحادي البعد (نموذج راش): (One-Parameter Logistic Model, Rasch Model)

يعد النموذج اللوغاريتمي ذو المعلمة الواحدة نموذج راش احد أوسع النماذج المستخدمة من بين نماذج نظرية استجابة الفقرة(Harris,1989).

ويطلق عليه بالنموذج اللوغاريتمي ذي المعلمة الواحدة؛ لأنه يعمل على تقدير معامل صعوبة الفقرة كما يطلق عليه اسم نموذج راش (Rasch model) نسبة إلى مطوره. ويعبر عن منحنى خصائص الفقرة للنموذج اللوغاريتمي ذو المعلمة الواحدة طبقا للمعادلة التالية:

$$Pi(\theta) = \frac{e^{(\theta - bi)}}{1 + e^{(\theta - bi)}} \qquad (1)$$

يشير Pi(θ) الى احتمالية أن يختار المفحوص عشوائيا مع قدرة مقدارها (θ) عندما يجيب على الفقرة (i) اجابة صحيحة.

bi هـو معلـم صـعوبة الفقـرة.

n عـدد فقـرات الاختبـار.

e هو العدد النيبري وقيمته (2.718).

معلمة (bi) للفقرة هي نقطة على متصل القدرة،حيث تكون احتمالية الإجابة الصحيحة عندها = 0.5 (القيمة العظمى لمعلمة bi)، اما القيمة

العظمى للقدرة (فتتحقق عندما يجيب المفحوص عن (50%) من الفقرات إجابة صحيحة).

عندما تحول قيم القدرة للمجموعة فان وسطها الحسابي يساوي (0) وانحرافها المعياري يساوي (1) وقد تأخذ bi قيما مختلفة ضمن المدى (2&2-)، فقيم bi التي تقترب من (2-) تعبر عن الفقرات السهلة جدا في حين أن قيم (2) تعبر عن الفقرات الصعبة جدا لمجموعة المفحوصين.

عليه نجد أن نموذج المعلمة الواحدة يعتمد على افتراضات محددة ويعتمد مدى مناسبة هذه الافتراضات على طبيعة البيانات. ولا يتطلب استخدام هذا النموذج القيام بعمليات رياضية معقدة بالإضافة إلى انه يمكن تقدير احتمالية هذا النموذج من خلال إجابة الفرد على الفقرة إجابة صحيحة وذلك من خلال معلمة واحدة هي معلمة صعوبة الفقرة فقط بصرف النظر عن حجم العينة وعدد الفقرات (Hambelton and Traub. 1971, 1973 . Wrigt , 1977).

ثانيا: النموذج اللوغاريتمي ذي المعلمتين:

يعتبر نموذج راش حالة خاصة من النموذج اللوغاريتمي ذي المعلمتين (Two-Parameter Logistic Model)، وهو النموذج الثاني ضمن هذه النظرية. ويعد لورد (1952) من أوائل العلماء الذين قاموا بتطوير نموذج استجابة الفقرة ذي المعلمتين معتمدا على التوزيع الطبيعي التراكمي.

وقد طور (Birnbaum, 1968) هذا النموذج اللوغاريتمي ذي المعلمتين وأصبح يعبر عنه بالمعادلة التالية:

$$Pi(\theta) = \frac{e^{Dai(\theta-bi)}}{1+e^{Dai(\theta-bi)}} \qquad (2)$$

$i=1,2,3,\ldots,n$

من الملاحظ ان معادلة (2)، هي نفس معادلة (1) المستخدمة في نموذج راش مع وجود اختلاف بسيط يتمثل بوجود عنصرين اضافيين هما: العامل (D) وهو عبارة عن عامل تقييس تم انتاجه لجعل الاقتران اللوغاريتمي قريبا قدر الامكان من الإقتران الطبيعي (Normal ogive function) أما العنصر الثاني الذي تم اضافته في نموذج ذي المعلمتين هو معلمة (ai) والذي يطلق عليه اسم معلمة تمييز الفقرة.

وعليه فإن هذا النموذج يفترض ان الفقرات تختلف في درجة صعوبتها وقدرتها التمييزية في ضوء المستويات المختلفة للقدرة.

أي انه يمكن لهذا النموذج تقدير احتمالية ان يجيب الفرد على الفقرة اجابة صحيحة من خلال معلمتين هما: معامل صعوبة الفقرة ومعامل تمييز الفقرة.

ثالثا: النموذج اللوغاريتمي ذي الثلاث معالم:

يعتبر النموذج اللوغاريتمي ذو الثلاث معالم (Three-Parameter Logistic Model) هو الأكثر عمومية من النموذجين السابقين، وذلك لأنه يضيف معلمة أخرى الى معالم الفقرة بالإضافة الى الصعوبة والتمييز وهي معلمة التخمين. ويعبر عنه رياضيا بالمعادلة التالية:

$$Pi(\theta) = c_i + (1-c_i)\frac{e^{Dai(\theta-bi)}}{1+e^{Dai(\theta-bi)}} \qquad (3)$$

$i= 1,2,3,\ldots,n$

حيث اننا عرفنا (D, ai, bi, pi) في المعادلات السابقة. أما المعلم الاضافي في هذا النموذج هو (ci) ويدعى بمعلمة تخمين الفقرة، وهذه المعلمة تمثل احتمالية ان يجيب المفحوص ذو القدرة المتدنية على الفقرة اجابة صحيحة.

وتدخل معلمة (ci) في النموذج وذلك لتأخذ بعين الاعتبار الاداء على نهاية الطرف الادنى من متصل القدرة، حيث ان التخمين يعتبر من أحد العوامل التي تتدخل في الإجابة عن فقرات الاختبار. ومثل هذه القيمة تفترض قيم تكون اقل من القيمة التي يمكن الحصول عليها من قبل المفحوص اذا خمن تلك الفقرة عشوائيا. وبهذا يكون النموذجين السابقين حالات خاصة من هذا النموذج.

والسؤال المطروح هنا: هو اي من هذه النماذج نختار؟

هناك وجهتا نظر حول ذلك، هما:

1- بعض الباحثين يفضلون اختيار النموذج أولا ثم في ضوئه يتم إختيار فقرات الاختبار لتطابق النموذج المختار.

2- البعض الآخر يفضل بناء الاختبار اولا ثم تحديد النموذج المطابق.

ولكن يرى هامبلتون ان هذا الامر متروك للمستخدم، ولكن المهم هو ان يلتزم بافتراضات النموذج وكيف يمكن لهذا النموذج ان يفسر نتائج الاختبار. كما تلعب عوامل أخرى في إختيار النموذج ترجع الى عدد المفحوصين و عدد فقرات الاختبار(الدرابيع، 2003).

وبعد أن تحدثنا عن النماذج الشائعة في نظرية الاستجابة على الفقرة فاننا سنقوم بعمل مقارنة بين هذه النماذج. تحاول نظرية استجابة الفقرة نمذجة العلاقة بين متغير غير ملاحظ وقدرة المفحوص واحتمالية ان يجيب المفحوص على اي فقرة من فقرات الاختبار. وتفترض هذه النظرية

بنماذجها الشائعة جميعها وجود قدرة كامنة منفردة للمفحوصين عادة تكون متصلة ومتغيرة وليس لها حدود وتميز بقدرة الفرد (θ).

جميع هذه النماذج لها معلمة صعوبة الفقرة (b) وهي نقطة انعكاس على متصل القدرة (θ) في النموذج ذي المعلمة الواحدة وذي المعلمتين وذلك عندما تكون احتمالية الاجابة الصحيحة على الفقرة (50%). بينما في النموذج ذي الثلاث معالم تكون احتمالية الاجابة الصحيحة على الفقرة = $(1+c_i)/2$.

نظريا فإن مدى قيم الصعوبة يتراوح من (∞^+-∞^-) في حين أن مدى قيم الصعوبة على المستوى العملي يتراوح من (3^+-3^-) وذلك عندما يكون وسطها الحسابي (=0) وانحرافها المعياري (=1). تكون الفقرات ذات قيم الصعوبة المرتفعة صعبة على الأفراد ذوي القدرة المنخفضة وتكون احتمالية الاجابة الصحيحة عن الفقرة منخفضة، أما الفقرات ذات قيم الصعوبة المنخفضة تكون فقرات سهلة على المفحوصين ذوي القدرة المرتفعة وتكون احتمالية الإجابة الصحيحة عن الفقرة مرتفعة.

تشتمل نماذج المعلمتين والثلاث معالم معلمة التمييز (a) التي تسمح للفقرات التفريق بين المفحوصين من حيث قدراتهم الحقيقية، وعمليا فأن معلمة التمييز (a) تعرف بانها ميل منحنى خصائص الفقرة عند نقطة الانعكاس (Baker, 1985)، تتراوح هذه القيمة من (∞^+-∞^-) مع قيم اقل من (0.2) لفقرات الاختيار من متعدد.

أما في النموذج اللوغاريتمي ذي الثلاث معالم فهناك معلمة خط التقاربي الادنى (c) التي تدعى عادة بمعلمة التخمين، و هذه المعلمة تسمح للمفحوصين (حتى ذوي القدرة المنخفضة) ان يجيبوا عن الفقرة اجابة صحيحة حتى لو كانت الفقرة متوسطة او صعبة، ومثل هذه القيمة تتراوح

نظريا بين (1-0) ولكنها قد تكون عمليا اقل من (0.3). وتستخدم النماذج السابقة جميعها الاقتران اللوغاريتمي لربط قدرة المفحوص ومعلمة الفقرة مع احتمالية اجابة الفرد عن الفقرة اجابة صحيحة. وكل فقرة يوجد لها ما يسمى بمنحنى خصائص الفقرة وهو ليس خطيا حيث ان التغير غير محدود ولكن الاحتمالية محدودة.

يعد تقدير المعالم هو اول اهم خطوة في تطبيق نظرية استجابة الفقرة على بيانات الاختبار والتي تصف نموذج استجابة الفرد عن الفقرة المنتقاة. وتعتمد احتمالية الاستجابة الصحيحة في نماذج استجابة الفقرة على قدرة المفحوص (θ) والمعلمة التي تصف الفقرة، ولكن إن كلا من القدرة ومعالم الفقرة مجهولة، اذن ما المعلوم؟ فالمعلوم هنا فقط هو استجابات المفحوصين الملاحظة عن فقرات الاختبار. ولكن المشكلة في التقدير هوتحديد قيمة (θ) لكل مفحوص وتحديد معالم الفقرة من الاستجابات عليها ان مثل هذه المشكلة شبيهة بالمشكلة التي تواجهنا في تحليل الانحدار المتعدد حيث ان المعالم التي تصف نموذج الانحدار تصفه من خلال الاستجابات الملاحظة على المتغير الذي لا يمكن تحديدها بشكل دقيق وانما يجري تقديرها.

هناك فرقان رئيسيان بين نموذج الانحدار ونماذج الاستجابة عن الفقرة ويتمثل الفرق الاول في أن نموذج الانحدار يعتبر نموذجا خطيا بينما نماذج الاستجابة ليست خطية أما الفرق الثاني والاهم هو ان المتغير المستقل في تحليل الانحدار هو متغير ملاحظ؛ اي ان العلامات على هذا المتغير يمكن ملاحظتها، اما في نماذج الاستجابة على الفقرة فإن المتغير المستقل (θ) هو غير ملاحظ. ولكن من الممكن عمل تقدير للمعالم من

خلال عدة طرق عندما تكون البيانات مطابقة تماما للنموذج (وهذا حدث غير طبيعي).

أما في الحالة الاخرى أي عندما تكون البيانات غير مطابقة للنموذج فإن استراتيجيتنا هي ان نجد قيم المعلمة التي تكون مطابقة، اي بمعنى اننا في تطبيقاتنا الفعلية فان كلا من معلمة الفقرة ومعالم المفحوص القدرات (θs) يجب تقديرها باستخدام استجابات المفحوصين على الفقرات، ولذلك بداية سنصف أول طريقة من طرق تقدير معالم القدرة وهي طريقة الارجحية القصوى (Maximum likelihood) لتقدير القدرة عندما تكون معالم الفقرة معلومة ثم سنصف طرق تقدير معالم الفقرة.

تقدير القدرة: (Estimation of Ability)

لنفرض أنه تم عشوائيا إختيار اجابات مفحوص لمجموعة (n) من الفقرات وفق نموذج الاجابة ($U_1, U_2,...,U_j, U_n$)، حيث ان (Uj) اما صفر(اجابة خاطئة) أو واحد (اجابة صحيحة) على الفقرة (j). وعليه فأن الإحتمالية المترابطة من نموذج الإستجابات الملاحظة هي ناتج ضرب الإحتماليات الملاحظة لكل استجابة على الفقرة (وذلك بواسطة افتراض الإستقلال المحلي) بحيث يؤخذ هذا بعين الاعتبار لصياغة الاقتران على الصورة التالية:

$$P(U_1,U_2,...,U_j,...,U_n/\theta)=P(U_1/\theta)\ P(U_2/\theta)...P(U_j/\theta)...P(U_n/\theta)$$

وباختصار أكثر:

$$P(U_1, U_2,..., U_j,..., U_n/\theta) = \prod_{j=1}^{n} P(U_j/\theta)$$

وكذلك:

$$P(U_1, U_2,..., U_j,..., U_n/\theta) = \prod_{j=1}^{n} P(U_j/\theta)^{u_j}[1-P(u_j/\theta)]^{1-u_j}$$

حيث أن:

$$Q_j=1-(P(U_j/\theta)) \quad , \quad P(U_j/\theta)=P_j \qquad (4)$$

وتعد هذه المعادلة تعبير للاحتمالية المترابطة لنمط الاستجابة. أما عندما يكون نمط الاستجابة ملاحظا $(U_j=u_j)$ فان التفسير الاحتمالي للإقتران لا يعد مناسبا، وعندها يدعى تعبير الاحتمالية المترابطة باقتران الأرجحية القصوى ويرمز له بالرمز $L(u_1,u_2,...,u_j,...,u_n/\theta)$

حيث أن (uj) هي الإستجابة الملاحظة على الفقرة (j) ، بحيث تصبح المعادلة على النحو التالي:

$$L(u_1, u_2,\dots, u_j,\dots, u_n/\theta) = \prod_{j=1}^{n} P_j^{\,uj}\, Qj^{(1-uj)} \qquad (5)$$

حيث أن $(Qj\&Pj)$ هما اقترانان للقدرة ومعالم الفقرات، فإن اقتران الارجحية القصوى هو اقتران من هذه المعالم، ولكن يعد افضل تقييس لاقتران الارجحية القصوى من خلال تحويل المعادلة باستخدام اللوغاريتمات فتصبح على النحو التالي:

$$\ln L(U/\theta) = \sum_{j=1}^{n}[u_j \ln P_j + (1-u_j)\ln(1-P_j)] \qquad (6)$$

هنا u هو متجه الاستجابات على الفقرة.

ان قيمة (θ) التي تجعل اقتران الارجحية القصوى للمفحوص يصل الى النهاية العظمى هو مايعرف على انه تقدير الارجحية القصوى ب (θ) لذلك المفحوص (maximum likelihood estimate). ان مشكلة ايجاد قيمة قصوى للاقتران هي ليست بالمشكلة البسيطة، لذلك فبالإمكان ايجاد قيمة قصوى للاقتران بواسطة طرق بحث بواسطة الحاسوب كما أن هناك طرقا اكثر

فاعلية تستخدم الحقيقة التي تقول أنه عندما يصل الاقتران الى قيمته القصوى، فان ميل الاقتران يساوي صفرا. وبناءا على ذلك فان تقدير الأرجحية القصوى يمكن تحديده بحل المعادلة الناتجة من المشتقة الاولى من اقتران الارجحية القصوى والعمل على جعلها تساوي صفرا.

ولكن مرة اخرى هذه المعادلة لا يمكن حلها مباشرة إذ يجب استخدام طرق تقريبية، ومن اكثر الطرق شيوعا هي طريقة نيوتن- رافسون.

ولسوء الحظ يمكن ان لايكون لاقتران الأرجحية قيم نهائية بحيث تكون عظمى وهذا يحدث عندما يجيب المفحوص عن كل الفقرات اجابة صحيحة أو يجيب عنها اجابة خاطئة في هذه الحالة فإن تقديرالأرجحية القصوى اما ($\theta=\infty$ أو $\theta=^{-}\infty$) في هذه الحالة يتم شطب جميع الحالات هذه لأنها لا تنسجم مع النموذج.

في بعض الاجابات الخاصة التي لا يمكن ان تكون مميزة كما في البديهيات فهذه لا يكون لها قيمة قصوى مطلقة في اقتران الأرجحية، وفي بعض الحالات يكون نمط الاستجابة مضللا مثل ان يجيب المفحوص عن فقرات صعبة ومميزة اجابة صحيحة ويجيب على الفقرات السهلة اجابة غير صحيحة في مثل هذه الحالة فان الطرق العددية المستخدمة لايجاد قيم قصوى سوف تنحرف، وغالبا ماتحدث هذه المشكلة في النموذج اللوغاريتمي ذي الثلاث معالم وليس مع النموذج احادي او ثنائي المعالم (Hambleton&Swamiathen, 1985) وبالإمكان حدوث هذه المشكلة في الاختبارات التي تحوي اكثر من اربعين فقرة.

ويمكن التغلب على مشكلة عدم تقدير اقصى للاقتران في بعض المواقف باستخدام طريقة (Bayesian estimation) وفكرتها الاساسية هي

تعديل اقتران الارجحية القصوى لدمجه مع اي من المعلومات السابقة التي لدينا عن معالم القدرة.

تقدير معالم الفقرة :

عند وصف طرق تقدير القدرة (θ) افترضنا أن معالم الفقرة معلومة ولكن عند بعض المواقف يجب مواجهة حقيقة ان معالم الفقرة يجب ان تقدر ايضا، ولتقدير هذه القدرة عندما تكون معالم الفقرة معلومة ستنشأ العديد من الفقرات للمفحوص ثم الحصول على اقتران الأرجحية لاستجابات المفحوص لعدد (n) من الفقرات. على العكس اذا اردنا ان نقدر معالم الفقرة عندما تكون القدرة معلومة ثم نحصل على اقتران الأرجحية لاستجابات (N) من المفحوصين على الفقرة وهو:

$$L(U_1, U_2, \ldots, U_n / \theta_{a,b,c}) = \prod_{i=1}^{N} Pi^{ui} Qi^{1-ui} \qquad (7)$$

حيث (a,b,c) هي معالم الفقرة (بالافتراض أنه نموذج ذي ثلاث معالم).

وهناك فروق بين اقترانات الأرجحية للمفحوص واقترانات الأرجحية لمعالم الفقرة، وهي باننا في اقتران الأرجحية لمعالم الفقرة لا نشترط حدوث افتراض الاستقلال المحلي، وكذلك تكون اقترانات الأرجحية لمعالم الفقرة متعددة الابعاد اي انه اقتران من ثلاث معالم، وبناء على ذلك فإنه لإيجاد إقتران الأرجحية القصوى للمعالم (a,b,c) علينا بداية ايجاد قيم لهذه المعالم تعبر عن اقصى قيمة لها في ثلاثة ابعاد وهذا يتم بايجاد المشتقة الاولى لاقتران الأرجحية القصوى بالنسبة لكل معلم من المعالم مع جعل هذه المشتقات تقترب من الصفر. اذن عندما تكون القدرة معلومة

لكل مفحوص فإننا بالإمكان إعتبار كل فقرة منفصلة بدون الرجوع الى فقرات اخرى. وبناء على ذلك فإن عملية التقدير يجب ان تكرر (n) من المرات وذلك لكل فقرة.

هناك طريقة أخرى لتقدير معالم الفقرة وهي (Maximum Marginal Likelihood Approach, MML)، وطريقة التقدير (MML) في نظرية استجابة الفقرة يعود الى (Lieberman, 1970) ولقد طورت هذه الحسابات بواسطة

(Bock& Aitkin, 1981). وهذه الطريقة تركز على ازالة معالم المفحوص من التقدير وتقدر فقط معالم الفقرة. ان احتمالية (Marginal) او احتمالية الاستجابة الصحيحة التي يجيبها المفحوص الذي اختير عشوائيا من المجتمع مع قدرته G(θ):

$$\int P(\theta) dG(\theta)$$

اذا قمنا باختيار عينة (N) من المفحوصين فان اقتران (MML) للبيانات الملاحظة هو:

$$L_m(a,b,c/U) = \prod_{j=1}^{N} \int \prod_{i=1}^{n} P_i(\theta) Q_i(\theta)^{1-u_{ij}} dG(\theta) \qquad (8)$$

وقد قدرت معالم الفقرة بايجاد قيمة قصوى لاقتران (MML)، وبالإمكان ايضا تقدير القدرة (θ). وعمليا فان (MML) تفترض قاعدة لتوزيع القدرة (θ) في مجتمع المفحوصين، سواء كانت احتمالية الإجابات الصحيحة في نظرية استجابة الفقرة ذات قيمة معينة او كان النموذج المفترض لتوزيع (θ) في المجتمع غير صحيح، فإن الخصائص الاحصائية يمكن أن تفشل (Hambelton&Swamithan, 1991) وبناء على ذلك فان طريقة

(MML) تطبق في النماذج احادي وثنائي وثلاثي المعالم بحيث ان هذه الطريقة تقوم بعمل اقتران تكاملي لمعلم القدرة وتقدر معالم الفقرة.

ولكن هناك حالات لا تكون فيها قدرة الفرد او معالم الفقرة معلومة وهذه الحالات شائعة جدا وتوجد مشكلة صعبة جدا في هذه الحالات وهي أننا يجب أن نعتبر أن استجابات جميع المفحوصين لجميع الفقرات هي استجابات تلقائية. ان اقتران (MEL) عندما يجيب (N) من المفحوصين لعدد (n) من الفقرات باستخدام افتراض الاستقلال المحلي هو:

$$L(u_1, u_2, \ldots, u_N / \theta_{a,b,c}) = \prod_{i=1}^{N} \prod_{j=1}^{n} P_{ij}^{u_{ij}} Q_{ij}^{1-u_{ij}} \qquad (9)$$

حيث (ui) استجابة المفحوص للفقرة (n)

(θ) متجه للقدرة (N)

(a, b, c) متجهات لمعالم الفقرة لاختبار مكون من (n) فقرة

(مع ملاحظة انه يجب تحقق افتراض الاستقلال المحلي حيث ان معالم القدرة ليست معلومة).

تسمى هذه الطريقة (JML The Joint Maximum Approach) ويعود أصل هذه الطريقة الى (Bimbaum, 1968).

وعلى الرغم من اتباع هذه الطريقة في كثير من التقديرات، الا ان لها مساوئ، ومنها :

1- تقديرات القدرة التامة وعلامات الصفر غير موجودة.

2- تقديرات معالم الفقرة للفقرة التي اجاب عليها جميع المفحوصين اجابة صحيحة او غير صحيحة ليست موجودة، وهذه حالات يجب حذفها و ذلك لعدم ملاءمتها للنموذج.

3- في النموذج ذي المعلمتين وذي الثلاث معالم طريقة (JML) للفقرة ومعالم القدرة لايمكن الحصول عليها حيث ان العالمان (Swaminathan & Gifford, 1979) عرضا عمليا انه بالإمكان الحصول على تقديرات لمعالم القدرة والفقرة اذا كان عدد المفحوصين كبير جدا وكذلك عدد الفقرات.

4- في النموذج ذي الثلاث معالم اذا لم توضع شروط محددة على قيم معالم الفقرة والقدرة فان الطريقة لايجاد تقديرات لها سوف تفشل.

تحدث هذه المشاكل في طريقة (JML) بسبب ان كلا من معالم الفقرة والقدرة تقدر تلقائيا ويمكن لهذه المشكلة ان تختفي اذا قدرت معالم الفقرة دون الرجوع الى معالم القدرة.

أما حديثا في تطبيقاتنا العملية تقدر هذه المعالم (الفقرة، القدرة) بواسطة برامج الحاسوب وذلك بسبب العدد الهائل من المعالم الذي يجب ان يقدر (اي لكل فقرة سيكون اما معلم او معلمين او ثلاث معالم ومعلم لكل مفحوص).

احد برامج التقدير الشائعة هو برنامج (LOGIST) (Wingersley,Barton,Lord, 1982) ويستخدم هذا البرنامج طريقة (JML) لتقدير معالم المفحوص والفقرة ويطبق هذا البرنامج على النماذج احادي المعلم وثنائي المعالم وثلاثي المعالم، وفي هذا البرنامج يرمز للبيانات صحيح، خطأ، محذوف، لايمكن الوصول اليه ويمكن استخدام طريقة (Newton- Raphson).

ان برنامج (LOGIST) فيه بعض القيود على مدى معالم (a,c) ومتصل (θ) لذلك يعتبر الوسط الحسابي مساويا الصفر، و الانحراف المعياري مساويا واحد صحيح. ويجب وضع كل مجموعة من المفحوصين والمعالم على نفس المتصل قبل مقارنة القدرات او معالم الفقرات في برامج الحاسوب (Hambleton & Swaminathem, 1985). ولكن هناك عيب في هذه الطريقة بانها لا تستطيع تزويدنا بتقديرالمعالم للمفحوص الذي اخذ صفر او علامة تامة او للفقرات التي اجاب عليها جميع المفخوصين اجابة صحيحة او لم يجب عليها احد.

تعتمد الدقة في تقدير المعالم علىعدد هذه المعالم والنموذج المختار وعدد المعالم التي سنقوم بتقديرها وعدد الفقرات والمفحوصين. ويرى (Hambleton,1983) أنه يجب استخدام عدد كبير من المفحوصين والفقرات عند تنفيذ برنامج (LOGIST) حيث حددها باكثر من (1000) مفحوص واكثر من (40) فقرة.

وهناك برنامج تقدير احدث وهو برنامج (BILOG) (Mislery&Bock,1984) وكما في برنامج (LOGIST) فبرنامج (BILOG) بإمكانه التعامل أيضا مع النماذج اللوغاريتمية الثلاثة (احادي وثنائي وثلاثي المعالم) ولكن هذا البرنامج يستخدم طريقة (Marginal Maximum Likelihood) بدلا من (JML) كما في برنامج (LOGIST).

وهناك ايضا برنامج (BICAL) وهذا البرنامج يطبق فقط على النموذج اللوغاريتمي احادي المعلم (Rasch model) بحيث يطبق في البرنامج تقدير معلمة واحدة (wright&stone, 1979) وهناك ايضا العديد من برامج الحاسوب التي عملت على تقدير المعالم. وفي هذه الدراسة فقد تم استخدام برنامج (BILOG)، وسنتناول الآن الحديث عن برنامج (BILOG)

بشيء من التفصيل وذلك لأهميته، يحصل برنامج (BILOG) على البيانات من خلال الإستجابات الملاحظة للفرد، ويستخدم هذا البرنامج في النموذج اللوغاريتمي ذي الثلاث معالم، بحيث يوجد لكل معلمة توزيعا مختلفا عن الآخر؛ حيث يستخدم هذا البرنامج التوزيع الطبيعي لمعلمة صعوبة الفقرة والتوزيع اللوغاريتمي لمعلمة تمييز الفقرة وتوزيع بيتا لمعلمة تخمين الفقرة. ويمكن تحديد هذه المعالم إما من قبل مستخدم البرنامج نفسه، أو من خلال البيانات التي حصل عليها.

تعريف الاختبار

تحدثنا فيما سبق عن نظرية استجابة الفقرة ونماذجها وطرق التقدير لهذه النماذج، ولكن لايمكن ان نطبق هذه النظرية الا عندما نبني اختبارا، اذن ما هو الاختبار، وما هي خطوات بناء الاختبار، وما هي انواعه؟ هذا ما سنحاول الاجابة عنه.

تتكرر كلمة اختبار او امتحان اكثر من غيرها في الاوساط التربوية؛ ويعود السبب في ذلك الى ان الاختبار من اكثر ادوات القياس والتقويم استخداما من قبل المدرسين في كافة المراحل (عودة ، 1993). بمعنى ان الاختبار هو احد المقاييس المستخدمة وبالتالي فان كلمة مقياس اعم واشمل من كلمة اختبار (test) أو امتحان (exam). وبالتالي يعرف الاختبار على أنه اداة قياس يتم اعدادها وفق طريقة منظمة من عدة خطوات تتضمن مجموعة من الاجراءات التي تخضع لشروط وقواعد محددة بغرض تحديد درجة امتلاك الفرد لسمة او قدرة معينة من خلال اجاباته عن عينة من المثيرات التي تمثل السمة او القدرة المرغوب قياسها (عودة ، 1993).

وقياسا على ذلك يمكن تعريف الاختبار التحصيلي: بأنه طريقة منظمة لتحديد مستوى تحصيل الطالب لمعلومات ومهارات في مادة دراسية كان قد تم تعلمها مسبقا بصفة رسمية من خلال اجاباته على عينة من الاسئلة (الفقرات) التي تمثل محتوى المادة الدراسية.

ولكن قبل وضع اي اختبار يجب اتباع خطوات معينة لبناء الاختبار التحصيلي وهي:

1- تحديد الغرض من الاختبار (الهدف الذي تريد قياسه).

2- اعداد جدول المواصفات حيث يشتمل جدول المواصفات على:

أ- تقسيم المادة الدراسية الى موضوعات او عناوين رئيسية وكما يمكن اعادة تقسيم هذه الموضوعات الى موضوعات اوعناوين فرعية.

ب- تحديد المجالات (معرفي/انفعالي/حركي) والمستويات التي يمكن تمييزها ضمن كل مجال تقع فيه الاهداف.

ج- تحديد وزن واهمية كل موضوع بالنسبة الى الموضوعات الاخرى في المادة الدراسية.

د- تحديد وزن واهمية كل مجال او المستوى في المجال، ويتحكم في هذا عدد الاهداف في كل مستوى.

ه- تحديد وزن واهمية كل موضوع وذلك بضرب النسبة المئوية للموضوع في النسبة المئوية للمستوى.

و- تحديد طول الاختبار (اي عدد فقراته).

3- صياغة الفقرات بعد اختيار الفئة المناسبة.

4- إخراج كراسة الاختبار وتتضمن التعليمات وترتيب الفقرات ترتيبا منطقيا.

5- التهيئة لتطبيق الاختبار وما يرافق ذلك من تنظيم الطلبة في القاعات وتوزيع الاوراق ...

تصنف الفقرات حسب نوع الفقرة (Item-Format) فهناك عدة اشكال من الفقر ات، أهمها:

1- المطابقة أو المزواجة (Matching): وهذا النوع على شكل جدول بعمودين أحدهما عمود المقدمات والثاني عمود الإجابات، ويطلب من الفرد أن يقوم بإختيار الإجابة المناسبة للمقدمة وذلك حسب التعليمات المعطاة في السؤال.

2- التكميل (Completation): ويطلب من الفرد إكمال معنى جملة بوضع كلمة أو شبه جملة ، أو تكميل شكل ....

3- الاختيار من متعدد (Multiple choice).

إن شكل اختبار الاختيار من متعدد ظهر بشكل سائد في الإجابات المنتقاة، ويستخدم معظم الممتحنيين أحد الأشكال الخمسة التالية، وهي:

1- الاختيار من متعدد المركب.

2- الاختيار البديل.

3- صح–خطأ.

4- صح – خطأ متعدد.

5- نظام الفقرات المعتمدة على المحتوى.

وهناك بعض جوانب القوة والضعف في الفقرات ذات الإجابة المنتقاة، ومن جوانب الضعف فيها:

1- قد يصل الفرد الى الإجابة بواسطة التخمين عندما لا تتوفر لديه المعلومات الصحيحة التي تمكنه من الإجابة.

2- إنها تقلل من الإبداع، أي لا تعط فرصة للفرد أن يعبر عن معرفته بلغته و مفرداته الخاصة.

3- تحتاج الى فترة زمنية طويلة نسبيا في اعدادها و كذلك تحتاج الى مهارة وخبرة لصياغتها.

أما بعض جوانب القوة تتمثل بما يلي:

1- تصحيح الإجابات بسهولة وبوقت أقصر.

2- توفر تغطية جيدة للمادة الدراسية.

3- لا تتأثر علامة الممتحن بذاتية المصحح، فالإجابة الصحيحة محددة وواضحة ولا تحتمل الخطأ (وهذا ما يؤدي الى تمتعها بدرجة عالية من الثبات).

إن فقرات الاختيار من متعدد هي الشكل الأوسع استخداما في اختبارات الإتجاهات والتحصيل، ولكن القضية التي تواجهنا عند بناءه هنا ماهي عدد البدائل (الخيارات) التي ممكن تضمينها لكل فقرة.

لقد حدد لورد (Lord, 1977) عدد الخيارات الموضوعية بأنه قد افترض أن زمن كل الاختبار (T) هو ثابت ومحدد مسبقا وهو يتناسب مع عدد الفقرات في الاختبار (n) وعدد الخيارات للفقرة (a) ثم افترض أن زمن الاختبار ممكن أن يعبر عنه كالتالي:

$$T=nat \qquad (10)$$

حيث (t) ثابت التناسب و(na) مجموع أرقام البدائل المثبتة في الاختبار.

ولقد عرض تافرسكي (Tversky,1964) ثلاثة معايير لتحديد عدد البدائل الممكنة، وهي:

1- عدد أنماط الإجابات الممكنة المميزة (القدرة التمييزية للاختبار).

2- قوة الاختبار (1- احتمال الحصول على علامة تامة بالصدفة فقط).

3- عدم التأكد من جميع أنماط الإستجابة المحتملة. وضمنها باستخدام أن (a=e=2.718) أو (a=3) بالتقريب.

ولكن الاقتران الذي أوجده تافرسكي غير مستخدم عادة في نظرية الاختبار، وفي الحقيقة هو لم يستخدم أي نموذج من نماذج القياس المحددة في حله، لكنه افترض أن جميع أنماط الإستجابات المحتملة غير متساوية تقريبا.

ويتألف اختبار الاختيار من متعدد من جزأين، الأول وهو الجذر ويشير الى العبارة التي تحدد المشكلة أو السؤال موضوع الإهتمام، والثاني وهو مجموعة البدائل، ويتراوح عددها عادة ما بين ثلاثة وخمسة بدائل بحيث تشير إحداها الى الجواب الصحيح أو الى أفضل الأجوبة الممكنة (نشواتي، 1991).

وتمتاز أسئلة الاختيار من متعدد بعدد من الخصائص تجعلها أداة قياس صادقة الى حد بعيد، فهي:

1- توفر الفرصة المناسبة لقياس أسلوب الطالب في تناول المشكلات ومعالجتها، وقدرته على التمييز بين الأحكام الصحيحة والخاطئة.
2- تؤدي الى ارتفاع مستوى ثبات النتائج بسبب انخفاض مستوى التخمين الناجم عن تعدد البدائل.
3- يمكن من تشخيص نواحي التفوق والضعف عند الطلاب.

وهناك العديد من الميزات الأخرى لاختبارات الاختيار من متعدد، ولكن يبقى السؤال المطروح هنا دائما ماهو العدد الأنسب للبدائل في اختبار الاختيار من متعدد؟ فبعض الباحثين وجد أن العدد الأمثل هو ثلاث بدائل مثلما توصلت اليه دراسة (Bruno&Dirkzwager,1995).

وهناك العديد من الدراسات التي أوصت بإستخدام أربعة بدائل عند وضع اختبار الاختيار من متعدد، وفي دراسات أخرى وجد أن العدد الأنسب هو خمسة بدائل.

وبناء على ما سبق نجد أن هناك تفاوتا في الأراء حول عدد البدائل الممكنة للفقرة، ومن هنا سنتناول في هذه الدراسة اختبارات الاختيار من متعدد بحيث يكون الاختبار مصمما بعدد بدائل مختلف (3,4,5)، ثم تطبيق النموذج اللوغاريتمي ذي الثلاث معالم في تقدير قدرة الفرد ومعالم الفقرات في كل مرة.

الدراسات السابقة

لقد اهتمت العديد من الدراسات السابقة بالبحث في النظرية الحديثة في القياس ومقارنتها مع النظرية التقليدية في القياس، واهتمت بعض الدراسات أيضا بالمقارنة بين نماذج نظرية استجابة الفقرة المختلفة وسوف نتناول بعض هذه الدراسات:

ففي دراسة أجراها أوين وفرومان (Owen and Froman, 1987) هدفت الى تجربة فعالية اختبار الاختيار من متعدد ذي الثلاث بدائل في اختبارات التحصيل. وذلك ببناء اختبارات متوازية من اختبارات ذات ثلاث بدائل وخمسة بدائل للفقرة وتوزيعها عشوائيا على طلبة كلية ما، وأشارت النتائج بعدم وجود فروق في معامل صعوبة الفقرة و معامل تمييز الفقرة والعلامة الكلية للاختبار، ولكن كان أقل وقت مستهلك هو في الاختبار ذي الثلاث بدائل.

وفي دراسة أجراها ستراتون وكاتس (Straton & Catts, 1980) هدفت لفحص خصائص اختبار الاختيار من متعدد باختلاف عدد البدائل لكل فقرة

(مرة بثلاث بدائل ومرة بأربعة بدائل) مع الاختبار ككل، أشارت النتائج بأن عدد البدائل للفقرة يرتبط ارتباطا عكسيا مع معامل صعوبة الفقرة ولكنه يرتبط ارتباطا طرديا مع معامل تمييز الفقرة، ولقد أوصيا باستخدام الاختبار من متعدد ذو الثلاث بدائل في غرفة الصف المثالية.

كما أجرى هامبلتون وتروب (Hambelton and Traub, 19971) دراسة هدفت الى مقارنة النموذج اللوغاريتمي ذي المعلمة الواحدة مع النموذج ذي المعلمتين والنموذج ذي الثلاث معالم وذلك باستخدام بيانات تم توليدها بطريقة المحاكاة، وقد تمت المقارنة بالإعتماد على منحنى خصائص الفقرة لتحديد دقة التقدير، حيث أشارت النتائج الى أن النموذج ذا الثلاث معالم كان الأفضل في تقدير قدرة الفرد وذلك لإحتوائه على معلمة التخمين وخاصة مع الأفراد ذوي القدرات المنخفضة.

وفي دراسة أجرتها (Anna Kubiak, 1986) هدفت الى تقدير كل من معالم الفقرة ومعلمة القدرة بواسطة النماذج الثلاث لنظرية استجابة الفقرة وذلك باستخدام بيانات تم توليدها بطريقة المحاكاة، حيث أشارت النتائج الى أن حجم العينة وطول الاختبار يعتمد على نموذج الإستجابة وتوزيع القدرة، و أشارت الى أن النموذج ذي الثلاث معالم يحتاج الى عينة كبيرة من الأفراد والفقرات بالمقارنة مع النماذج أحادي و ثنائي المعالم.

وفي دراسة قام بها (Seock-Ho KIM, 1991) هدفت الى اشتقاق ومن ثم المقارنة بين طريقة التقدير (Joint maximum likelihood) وطريقة (Marginal Bayesian) وذلك لتقدير معالم الفقرة والقدرة في نظرية الإستجابة على الفقرة، ولقد كانت الدراسة مكونة من جزئين رئيسيين وهما: الإشتقاقات الرياضية، والمقارنة التطبيقية. وأشارت النتائج الى أن

طريقة (Marginal Bayesian) أفضل من طريقة التقدير المقارنة معها لتقدير معالم الفقرة.

ولقد قام فان (Fan,1998) بدراسة هدفت الى مقارنة الى أي مدى تختلف النظرية التقليدية عن نظرية استجابة الفقرة في:

1- تقدير قدرات المفحوصين.
2- تقدير معاملات صعوبة الفقرة.
3- تقدير معاملات تمييز الفقرة.

ولقد جرى تحليل نتائج الاختبار حسب النظرية التقليدية ونظرية استجابة الفقرة بنماذجها الثلاث باستخدام برنامج (BILOG) واظهرت النتائج بأن تقدير قدرات الأفراد من خلال النظرية التقليدية كانت مشابهة الى حد كبير مع التقديرات من خلال نظرية استجابة الفقرة بنماذجها الثلاث باستخدام معاملات الإرتباط، وأشارت النتائج الى أن معاملات صعوبة الفقرة من خلال النظرية التقليدية تتشابه مع التقديرات من خلال نظرية استجابة الفقرة باستخدام معاملات الإرتباط، و أن معاملات تمييز الفقرات من خلال النظرية التقليدية تتشابه بصورة أقل من تقديرات نظرية استجابة الفقرة باستخدام معاملات الإرتباط.

وفي دراسة قام بها (Hulin, Lissak and Drasgow, 1982) هدفت الى دراسة أثر حجم العينة وطول الاختبار على دقة تقدير قدرة الفرد ومعالم الفقرة باستخدام النموذج اللوغاريتمي ذي المعلمتين وذي الثلاثة معالم، حيث تم استخدام معامل الإرتباط بين قدرة الفرد الحقيقية والمقدرة، وبين معالم الفقرات الحقيقية والمقدرة واستخدام الجذر التربيعي لمتوسط مربعات الإنحرافات كمعايير لتقييم دقة التقدير حيث تم استخدام أربعة مستويات لحجم العينة وهي (200,500,1000,2000) فردا وثلاثة مستويات لطول الاختبار

(15,30,60) فقرة. حيث أشارت النتائج فيما يتعلق بالنموذج اللوغاريتمي ذي الثلاثة معالم الى أن معامل الإرتباط بين المعامل الحقيقي لتمييز الفقرة والمعامل المقدر كان (0.36) عند استخدام اختبار مكون من (15) فقرة وحجم عينة مكون من (200) فرد وكان معامل الإرتباط يساوي (0.84) عند استخدام اختبار مكون من (60) فقرة وحجم عينة مكون من (2000) فرد. كما وجد أن معامل الإرتباط بين المعامل الحقيقي لصعوبة الفقرة والمعامل المقدر يساوي (0.72) عند استخدام الاختبار المكون من (15) فقرة وحجم العينة المكون من (200) فرد، في حين ازداد هذا المعامل ليصبح (0.97) عند استخدام الاختبار المكون من (60) فقرة وحجم العينة المكون من (2000) فرد. وبشكل عام تشير هذه النتائج الى أنه بزيادة حجم العينة وطول الاختبار فإن دقة التقدير لقدرة الفرد و لمعالم الفقرة يزداد.

ويلاحظ من هذه الدراسات أن بعضها قد أشار الى إستخدام الاختبار ذي الثلاث بدائل؛ وذلك لأن عدد البدائل للفقرة يرتبط عكسيا مع معامل صعوبة الفقرة وطرديا مع معامل تمييز الفقرة، في حين أشارت بعض الدراسات الأخرى الى عدم وجود فروق في معامل صعوبة وتمييز الفقرة مع العلامة الكلية للاختبار.

في حين أشارت بعض الدراسات الى أن النموذج اللوغاريتمي ذي الثلاث معالم هو الأفضل في تقدير قدرة وذلك لإحتوائه على معلمة التخمين.

لذلك فإن هذه الدراسة تحاول تطبيق النموذج اللوغاريتمي ذي الثلاث معالم في تقدير قدرة الفرد و معالم الفقرات في اختبار الاختيار من متعدد يعزى الى اختلاف عدد البدائل (3،4،5).

وفي دراسة أجراها هاريز (Harris, 1989) للمقارنة بين النماذج الثلاثة في النظرية الحديثة في القياس، بحيث تم توضيح المعادلة الرياضية

لكل نموذج، وتمت المقارنة بين هذه النماذج الثلاثة وتضمنت الدراسة الرسوم البيانية لتوضيح تأثير التغيرات في المعالم الثلاثة، وهي: معلمة التمييز، معلمة الصعوبة، معلمة التخمين، وكذلك استخدمت بيانات موحدة ل(3000) مفحوص لتوضيح التأثيرات في تقدير المعالم ومقارنة تقديرات هذه المعالم عند تطبيق نماذج مختلفة. ولم تشر الدراسة الى تفوق أي من النماذج على الآخر إلا أنه يجب استخدام النموذج الأكثر انسجاما مع بيانات العينة المستخدمة لأنه سوف يعطي تقديرات أدق لمعالم الفقرات من غيره من النماذج الأخرى. كما وبينت الدراسة كيفية مناقشة افتراضات نماذج النظرية الحديثة مثل أحادية البعد والاستقلال المحلي.

التعريفات الإجرائية:

تم تعريف المصطلحات التي وردت في هذه الدراسة على النحو التالي:

صعوبة الفقرة:

تشير صعوبة الفقرة وفق النظرية التقليدية في القياس الى نسبة الإستجابات الصحيحة. وتشير صعوبة الفقرة وفق النظرية الحديثة في القياس الى القيمة التي تتدرج على متصل القدرة والتي يكون احتمال اجابة الفقرة عندها اجابة صحيحة مساويا (50 %).

تمييز الفقرة:

يشيرتمييز الفقرة وفق النظرية التقليدية الى الفرق بين نسبة الإستجابات الصحيحة بين الفئتين العليا والدنيا من المفحوصين، ويمكن حسابه ايضا من خلال معامل الإرتباط المعدل (بوينت بايسيريال) لأنه

يسلك نفس سلوك المؤشر السابق. ويشير تمييز الفقرة وفق النظرية الحديثة الى ميل المماس لمنحنى خصائص الفقرة عند النقطة التي تحدد صعوبتـــها

قدرة الفرد:

تشير قدرة الفرد وفق النظرية التقليدية الى العلامة المقدرة التي يحصل عليها الفرد عند اجابته تحريريا أو شفويا على فقرات اختبار يقيس سمة معينة. وتشير قدرة الفرد وفق النظرية الحديثة في القياس الى اللوغاريتم الطبيعي لمرجح نجاح الفرد على الفقرات التي تعبر نقطة صفر التدريج عن صعوبتها (Wright and Stone , 1979).

التخمين:

هو احد معالم الفقرات ويعني احتمال حصول المفحوص ذي القدرة المتدنية على الإجابةالصحيحة بواسطة الصدفة.

متغيرات الدراسة:

المتغير المستقل: عدد البدائل في اختبار الاختيار من متعدد.

المتغيرات التابعة:

1- قدرة الفرد.
2- معامل صعوبة الفقرة.
3- معامل تمييز الفقرة.
4- معامل التخمين للأداء على الفقرة.

الفصل الثالث

المنهجية و الإجراءات

يتناول هذا الفصل وصفا لمجتمع الدراسة وعينتها، والأداة المستخدمة فيها والإجراءات التي تم اتباعها، وأسلوب معالجة البيانات إحصائيا.

مجتمع الدراسة:

تكون مجتمع الدراسة من طلبة الصف العاشر الأساسي التابعين لمدارس وزارة التربية والتعليم في تربية عمان الرابعة للعام الدراسي 2002/2001 والبالغ عددهم (10245) طالبا وطالبة.

عينة الدراسة:

تم اختيار عينة الدراسة وفق الطريقة العشوائية البسيطة على مستوى الصفوف والشعب في المدارس بما يتناسب مع توزيع الأفراد بالنسبة لمجتمع الدراسة، حيث تم اختيار (400) طالب وطالبة لاختبار المكون من ثلاثة بدائل، و(400) طالب وطالبة لاختبار المكون من أربعة بدائل و(400) طالب وطالبة لاختبار المكون من خمسة بدائل على إفتراض التكافؤ بين المجموعات الثلاث من حيث الصف والمنهاج.

أداة الدراسة:

لجمع البيانات اللازمة للإجابة عن أسئلة الدراسة تم بناء اختبار تحصيلي في مادة الرياضيات (الفصل الأول) للصف العاشر وذلك وفق جدول المواصفات حتى نحقق صدق المحتوى للاختبار، ويشمل الاختبار

على (35) فقرة إختيار من متعدد، بحيث كانت نفس الأسئلة ولكن بتغيير عدد البدائل في كل مرة من مرات الاختبار فلقد كان مرة بثلاثة بدائل ثم طبق الاختبار على (400) طالب وطالبة، ومرة بأربعة بدائل ثم طبق على (400) طالب وطالبة آخرين، ومرة بخمسة بدائل ثم طبق على (400) طالب وطالبة آخرين.

صدق الاختبار:

لقد تم استقصاء صدق هذا الاختبار عن طريق عرضها على محكمين مختصين في القياس والتقويم وعلم النفس التربوي للحكم على مدى مناسبة فقرات المقياس(الاختبار)، ووضوح لغته، وقد أجمع (80%) من المحكمين على اعتماد فقرات الاختبار في حين تم إجراء بعض التعديلات على بعض الفقرات لتكون أكثر وضوحا.

ثبات الاختبار:

للتأكد من ثبات الاختبار، فقد تم حساب معامل الإتساق الداخلي باستخدام معادلة كرونباخ ألفا من خلال تطبيق الاختبار على عينة مكونة من (400) طالبا و طالبة من الصف العاشر، وقد بلغت قيمته (0.75)، وفي ضوء هذه النتيجة اعتبرت هذه الأداة مناسبة ومستوفية للخصائص السيكومترية التي تؤهلها لأغراض هذه الدراسة.

اجراءات تنفيذ الدراسة

بعد الحصول على الموافقة الرسمية من وزارة التربية والتعليم لتطبيق الاختبار على طلبة الصف العاشر في مدارس تربية عمان الرابعة، تم إعداد قائمة بأسماء المدارس الأساسية الموجودة في هذه

المنطقة، بعد ذلك تم الذهاب الى هذه المدارس، ومقابلة مدير المدرسة وسؤاله عن عدد طلبة الصف العاشر في مدرسته وهل يسمح بتطبيق الاختبار عليهم أم لا؟ وبعد ذلك تعريف المدير بأغراض هذه الدراسة وأهدافها. وإتباع نفس الأسلوب مع باقي المدارس، ثم إختيار عينة بالطريقة العشوائية البسيطة من هذه المدارس، مع مراعاة الأعداد الموجودة في كل مدرسة. ولضمان الحصول على جدية المستجيبين على فقرات هذا الاختبار لم يتم إعلامهم مطلقا بأن هذا اختبار لأغراض الدراسة وإنما هو اختبار فعلي من اختبارات المدرسة وسوف تحسب علامته وتضاف الى المعدل، وذلك بالإتفاق مع مدراء المدارس ومعلمي مادة الرياضيات.

المعالجة الإحصائية

بعد أن تم جمع جميع الاختبارات من المدارس، فرغت الإجابات بحيث اعتبرت كل إجابة صحيحة تساوي واحد، وإعتبار كل إجابة خاطئة تساوي صفر. ثم حللت هذه الإجابات باستخدام برنامج (BILOG) باستخدام الحاسوب، ثم تفريغ هذه التحليلات على برنامج (SPSS) وللإجابة عن أسئلة الدراسة تم استخدام تحليل التباين الأحادي.

الفصل الرابع

عرض النتائج

يتناول هذا الفصل عرضا للنتائج التي توصلت إليها هذه الدراسة والمرتبطة بأسئلتها التالية:

النتائج المتعلقة بالسؤال الأول:

هل هناك فروقا ذات دلالة احصائية في قدرة الفرد (الدرجة الكلية) على اختبار الاختيار من متعدد يعزى الى اختلاف عدد البدائل وفق النظرية التقليدية في القياس؟

وللإجابة عن هذا السؤال فقد تم اجراء تحليل التباين أحادي (one_ way ANOVA) في قدرة الفرد على اختبار الاختيار من متعدد، ويوضح جدول رقم (1) نتائج هذا التحليل.

جدول رقم (1)

نتائج تحليل التباين الأحادي لفحص دلالة الفروق في قدرة الفرد(الدرجة الكلية) باختلاف عدد البدائل وفق النظرية التقليدية في القياس

مصدر التباين	مجموع المربعات	درجات الحرية	متوسط مجموع المربعات	ف	مستوى الدلالة
بين المجموعات	1089.5	2	549.2	25.2	0.0
داخل المجموعات	26080.8	1194	21.8		
المجموع	27179.3	1196			

يلاحظ من نتائج تحليل التباين الأحادي كما في جدول رقم (1) الى وجود فروق دالة احصائيا في العلامة الكلية باختلاف عدد البدائل.

وللكشف عن مصادر الفروق في العلامة الكلية للاختبار تم استخدام اختبار توكي (Tukey HSD) للمقارنات البعدية ويبين الجدول رقم (2) نتائج هذه المقارنات.

جدول رقم (2)

نتائج المقارنات البعدية في العلامة الكلية وفق النظرية التقليدية في القياس

عدد البدائل	ثلاث بدائل	أربع بدائل	خمس بدائل
ثلاث بدائل	—	1.9821*-	2.0763*-
أربع بدائل	1.9821*	—	0.00942-
خمس بدائل	2.0763*	0.009419	—

يتبين من نتائج المقارنات البعدية كما في جدول رقم (2) وجود فروق ذات دلالة احصائية في العلامة الكلية للاختبار باختلاف عدد البدائل بين الاختبارات ذات الثلاث والأربع والخمس بدائل وذلك لصالح الاختبار ذي الخمس بدائل حيث بلغ متوسط الاختبار ذي الخمس بدائل (19.5) أكثر من متوسطات العلامة الكلية للاختبارات ذات الأربع والثلاث بدائل، حيث بلغ متوسط الاختبار ذي الثلاث بدائل (17.4) بينما بلغ متوسط الاختبار ذي الأربع بدائل(19.4).

النتائج المتعلقة بالسؤال الثاني :

هل هناك فروقا ذات دلالة احصائية في قدرة الفرد على اختبار الاختيار من متعدد باختلاف عدد البدائل وفق النظرية الحديثة في القياس؟

وللإجابة عن هذا السؤال فقد تم اجراء تحليل التباين الأحادي لقدرة الفرد ويبين الجدول رقم (3) نتائج تحليل التباين لفحص الفروق في قدرة الفرد تعزى الى اختلاف عدد البدائل في اختبار الاختيار من متعدد.

جدول رقم (3)

نتائج تحليل التباين الأحادي لفحص دلالة الفروق بين قدرة الفرد في اختبار الاختيار من متعدد باختلاف عدد البدائل وفق النظرية الحديثة في القياس

مصدر التباين	مجموع المربعات	درجات الحرية	متوسط مجموع المربعات	ف	مستوى الدلالة
بين المجموعات	0.207	2	0.103	0.705	0.839
داخل المجموعات	706.16	1195	0.591		
المجموع	706.37	1197			

يلاحظ من جدول رقم (3) عدم وجود فروق ذات دلالة إحصائية (α> 0.05) في قدرة الفرد يعزى الى اختلاف عدد البدائل.

النتائج المتعلقة بالسؤال الثالث:

هل هناك فروقا ذات دلالة احصائية في صعوبة الفقرة يعزى الى اختلاف عدد البدائل وفق النظرية التقليدية في القياس؟

يبين جدول رقم (4) نتائج تحليل التباين الأحادي لفحص دلالة الفروق في معلمة صعوبة الفقرة للفرد يعزى الى اختلاف عدد البدائل في اختبار الاختيار من متعدد.

جدول رقم(4)

نتائج تحليل التباين الأحادي لفحص دلالة الفروق في معلمة صعوبة الفقرة باختلاف عدد البدائل وفق النظرية التقليدية في القياس

مصدر التباين	مجموع المربعات	درجات الحرية	متوسط مجموع المربعات	ف	مستوى الدلالة
بين المجموعات	0.609	2	0.305	1.604	0.206
داخل المجموعات	19.376	102	0.190		
المجموع	19.986	104			

أشارت نتائج تحليل التباين الأحادي كما في جدول رقم (4) الى عدم وجود فروق دالة احصائيا في معلمة صعوبة الفقرة باختلاف عدد البدائل في اختبار الاختيار من متعدد عند مستوى دلالة (α<0.5).

النتائج المتعلقة بالسؤال الرابع:

هل هناك فروقا ذات دلالة احصائية في صعوبة الفقرة على اختبار الاختيار من متعدد يعزى الى اختلاف عدد البدائل وفق النظرية الحديثة في القياس؟

ويشير جدول رقم (5) الى نتائج تحليل التباين الأحادي لفحص دلالة الفروق في معلمة صعوبة الفقرة على اختبار الاختيار من متعدد يعزى الى اختلاف عدد البدائل.

جدول رقم (5)

نتائج تحليل التباين الأحادي لفحص دلالة الفروق في معلمة صعوبة الفقرة باختلاف عدد البدائل وفق النظرية الحديثة في القياس

مصدر التباين	مجموع المربعات	درجات الحرية	متوسط مجموع المربعات	ف	مستوى الدلالة
بين المجموعات	18.963	2	9.482	22.902	0.000
داخل المجموعات	42.228	102	0.414		
المجموع	61.191	104			

حيث يلاحظ من خلال جدول رقم (5) أن نتائج تحليل التباين الأحادي تشير الى وجود فروق ذات دلالة احصائية عند مستوى (α< 0.05) في معلمة الصعوبة باختلاف عدد البدائل في اختبار الإختيار من متعدد.

وللكشف عن مصادر الفروق في معلمة صعوبة الفقرة للإختبار باختلاف عدد البدائل في اختبار الإختيار من متعدد تم استخدام اختبار

توكي (Tukey HSD) للمقارنات البعدية، والجدول رقم (6) يبين نتائج هذه المقارنات.

جدول رقم (6)

نتائج المقارنات البعدية لمعلمة صعوبة الفقرة وفق النظرية الحديثة في القياس

عدد البدائل	ثلاثة بدائل	أربعة بدائل	خمسة بدائل
ثلاثة بدائل	–	0.86*	0.93*
أربعة بدائل	0.87-*	–	0.006
خمسة بدائل	0.86-*	0.006-	

يتبين من نتائج الإختبار البعدي كما في جدول رقم (6) أن هناك فروقا دالة احصائيا في معلمة الصعوبة باختلاف عدد البدائل عند مستوى دلالة ($0.05>\alpha$) بين الإختبار ذي الثلاث بدائل و الإختبارات ذات الأربع والخمس بدائل وذلك لصالح الإختبار ذي الثلاث بدائل حيث بلغ متوسط معلمة الصعوبة للإختبار ذي الثلاث بدائل (1.39) أكثر من متوسطات معلمة الصعوبة للإختبارات ذي الأربع و الخمس بدائل، حيث بلغ متوسط معلمة صعوبة الفقرة للإختبار ذي الأربع بدائل (0.52) بينما بلغ متوسط معلمة صعوبة الفقرة للإختبار ذي الخمس بدائل (0.45).

النتائج المتعلقة بالسؤال الخامس:

هل هناك فروقا ذات دلالة احصائية في تمييز الفقرة على إختبار الإختيار من متعدد بإختلاف عدد البدائل وفق النظرية التقليدية في القياس؟

و يبين جدول رقم (7) نتائج تحليل التباين الأحادي لفحص دلالة الفروق في معلمة تمييز الفقرة للفرد يعزى الى اختلاف عدد البدائل في اختبار الإختيار من متعدد.

جدول رقم (7)

نتائج تحليل التباين الأحادي لفحص دلالة الفروق في معلمة تمييز الفقرة باختلاف عدد البدائل وفق النظرية التقليدية في القياس

مصدر التباين	مجموع المربعات	درجات الحرية	متوسط مجموع المربعات	ف	مستوى الدلالة
بين المجموعات	0.170	2	0.009	24.23	0.000
داخل المجموعات	0.358	102	0.0004		
المجموع	0.528	104			

وكما يلاحظ في جدول رقم (7) نتائج تحليل التباين الأحادي تشير الى وجود فروق دالة احصائيا في معلمة تمييز الفقرات باختلاف عدد البدائل في اختبار الاختيار من متعدد عند مستوى دلالة (α>0.05) .

وللكشف عن مصادر الفروق في معلمة تمييز الفقرات يعزى الى اختلاف عدد البدائل في اختبار الإختيار من متعدد تم استخدام اختبار توكي (Tukey HSD) للمقارنات البعدية، ويبين جدول رقم (8) نتائج هذه المقارنات.

جدول رقم (8)

نتائج المقارنات البعدية لمعلمة تمييز الفقرة وفق النظرية التقليدية في القياس

عدد البدائل	ثلاثة بدائل	أربعة بدائل	خمسة بدائل
ثلاثة بدائل	—	-0.00811*	-0.00891*
أربعة بدائل	0.008109*	—	-0.00800
خمسة بدائل	0.0089009	0.008000	

أشارت نتائج تحليل المقارنات البعدية كما في جدول رقم (8) الى وجود فروق ذات دلالة احصائية في معلمة تمييز الفقرة يعزى الى اختلاف عدد البدائل بين الاختبارات ذات الثلاث والأربع بدائل وذلك لصالح الاختبار ذي الأربع بدائل حيث بلغ متوسط معلمة تمييز الفقرات للاختبار ذي الأربع بدائل (0.29) وهو أكثر من متوسط معلمة التمييز للاختبار ذي الثلاث بدائل (0.21). وكذلك بين الاختبارات ذات الثلاث والخمس بدائل وذلك لصالح الاختبار ذي الخمس بدائل، حيث بلغ متوسط معلمة تمييز الفقرة للاختبار ذي الخمس بدائل (0.28).

النتائج المتعلقة بالسؤال السادس:

هل هناك فروقا ذات دلالة احصائية في تمييز الفقرة على اختبار الاختيار من متعدد يعزى الى اختلاف عدد البدائل وفق النظرية الحديثة في القياس ؟

ويبين الجدول رقم (9) نتائج تحليل التباين الأحادي لفحص الفروق في معلمة تمييز الفقرة في اختبار الاختيار من متعدد يعزى الى اختلاف عدد البدائل.

جدول رقم (9)

نتائج تحليل التباين الأحادي لفحص دلالة الفروق في معلمة تمييز الفقرة باختلاف عدد البدائل وفق النظرية الحديثة في القياس

مصدر التباين	مجموع المربعات	درجات الحرية	متوسط مجموع المربعات	ف	مستوى الدلالة
بين المجموعات	1.501	2	0.75	3.104	0.049
داخل المجموعات	24.651	102	0.24		
المجموع	26.151	104			

أشارت نتائج تحليل التباين الأحادي كما هو موضح في جدول رقم (9) عن وجود فروق ذات دلالة احصائية في معلمة التمييز يعزى الى اختلاف عدد البدائل في اختبار الاختيار من متعدد. وللكشف عن مصادر الفروق المختلفة في معلمة تمييز الفقرة للاختبار يعزى الى اختلاف عدد البدائل تم استخدام اختبار توكي (Tukey HSD) للمقارنات البعدية والجدول رقم (10) يبين نتائج هذه المقارنات.

جدول رقم (10)

نتائج المقارنات البعدية لمعلمة تمييز الفقرة وفق النظرية الحديثة في القياس

عدد البدائل	ثلاثة بدائل	أربعة بدائل	خمسة بدائل
ثلاثة بدائل	—	0.1326-	0.2924-*
أربعة بدائل	0.1326	—	0.1598-
خمسة بدائل	0.2924*	0.1598	

يتبين من نتائج الاختبار البعدي كما في جدول رقم (10) أن هناك فروقا دالة احصائيا في معلمة تمييز الفقرة باختلاف عدد البدائل بين الاختبارات ذات الثلاث والخمس بدائل عند مستوى دلالة (α<0.05) وذلك لصالح الاختبار ذي الخمس بدائل حيث بلغ متوسط معلمة التميز للاختبار ذي الخمس بدائل (0.95) أكثر من متوسط معلمة التمييز للاختبار ذي الثلاث بدائل، حيث بلغ متوسط معلمة تمييز الفقرة للاختبار ذي الثلاث بدائل (0.66).

النتائج المتعلقة بالسؤال السابع:

هل هناك فروقا ذات دلالة احصائية في تخمين الفقرة على اختبار الاختيار من متعدد يعزى الى اختلاف عدد البدائل في النظرية الحديثة في القياس؟

يبين الجدول رقم (11) نتائج تحليل التباين الأحادي لفحص دلالة الفروق في معلمة التخمين في اختبار الاختيار من متعدد يعزى الى اختلاف عدد البدائل.

جدول رقم (11)

نتائج تحليل التباين الأحادي لفحص دلالة الفروق في معلمة تخمين الفقرة باختلاف عدد البدائل وفق النظرية الحديثة في القياس

مصدر التباين	مجموع المربعات	درجات الحرية	متوسط مجموع المربعات	ف	مستوى الدلالة
بين المجموعات	0.00492	2	0.002	7.521	0.001
داخل المجموعات	0.305	102	0.003		
المجموع	0.350	104			

يلاحظ من نتائج تحليل التباين الأحادي كما في جدول رقم (11) الى وجود فروق ذات دلالة احصائية في معلمة تخمين الفقرات يعزى الى اختلاف عدد البدائل في اختبار الاختيار من متعدد.

وللكشف عن مصادر الفروق في معلمة التخمين للاختبار باختلاف عدد البدائل تم استخدام اختبار توكي (Tukey HSD) للمقارنات البعدية والجدول رقم (12) نتائج هذه المقارنات.

جدول رقم (12)

نتائج المقارنات البعدية لمعلمة تخمين الفقرة وفق النظرية الحديثة في القياس

عدد البدائل	ثلاثة بدائل	أربعة بدائل	خمسة بدائل
ثلاثة بدائل	—	0.004563*	0.004189*
أربعة بدائل	0.00456-*	—	0.000374-
خمسة بدائل	0.00419-*	0.0003743	—

يتبين من نتائج الاختبار البعدي كما في جدول رقم (12) أن هناك فروقا دالة احصائيا باختلاف عدد البدائل بين الاختبارات ذات الثلاث والأربع والخمس بدائل عند مستوى دلالة ($0.05 > \alpha$) وذلك لصالح الاختبار ذي الثلاث بدائل حيث بلغ متوسط معلمة التخمين للاختبار ذي الثلاث بدائل (0.27) أكثر من متوسطات معلمة التخمين للاختبارات ذات الأربع والخمس بدائل، حيث بلغ متوسط معلمة التخمين للاختبار ذي الأربع بدائل (0.22) بينما بلغ متوسط معلمة التخمين للاختبار ذي الخمس بدائل (0.22).

الفصل الخامس

المناقشة و التوصيات

من أهم ما توصلت اليه هذه الدراسة عدم وجود الأثر المعنوي لدلالة الفروق في قدرة الفرد على اختبار الاختيار من متعدد باختلاف عدد البدائل وذلك وفق النظرية الحديثة في القياس، ويمكن عزو ذلك الى أن الاختبار مكون من عدد قليل من الفقرات، بحيث كانت هذه النتيجة لا تتفق مع الدراسات السابقة والتي أشارت نتائجها الى أنه في حالة إستخدام عدد كبير من الفقرات والأفراد فإن دقة التقدير لقدرة الفرد تزداد وبالتحديد عند استخدام النموذج اللوغاريتمي ذي الثلاث معالم كدراسة سوامنثان وجفورد، عام 1977 (Swaminthan and Gifford,1979) ودراسة هامبلتون وكوك (Hambleton and Cook, 1980).

في حين أظهرت النتائج وجود الأثر المعنوي لدلالة الفروق في قدرة الفرد على اختبار الاختيار من متعدد باختلاف عدد البدائل وفقا للنظرية الكلاسيكية في القياس، ولقد كان ذلك لصالح الاختبار المكون من خمسة بدائل، بسبب أن قيمة المتوسط للاختبار ذي الخمس بدائل كانت أكثر من المتوسط للاختبارات ذات الثلاث بدائل والأربعة بدائل.

وعند إجراء المقارنات البعدية، وجد أن الاختبار ذي الخمس بدائل هو الأفضل في تقدير قدرة الفرد. وقد يرجع ذلك الى أن زيادة عدد البدائل ينعكس كذلك على القدرة التمييزية للفقرة بين المستجيبين. وهذه النتائج لا تتفق مع دراسة بوديسكو ونيفو (Budesco & Nevo,1985).

إضافة الى ذلك، فقد أشارت نتائج هذه الدراسة الى نتيجة هامة بما يتعلق بوجود فروق ذات دلالة احصائية في صعوبة الفقرة على اختبار الاختيار من متعدد يعزى الى اختلاف عدد البدائل وفق النظرية الحديثة في

القياس، وذلك لصالح الاختبار ذي الثلاث بدائل، بسبب أن المتوسط للاختبار ذي الثلاث بدائل كان أكثر من المتوسط للاختبارات ذات الأربع و الخمس بدائل. وعند إجراء المقارنات البعدية، وجد أن الاختبار ذي الثلاث بدائل هو الأفضل في تقدير صعوبة الفقرة.

في حين أظهرت النتائج عن عدم وجود الأثر المعنوي لدلالة الفروق في صعوبة الفقرة على اختبار الاختيار من متعدد باختلاف عدد البدائل وفقا للنظرية الكلاسيكية في القياس، وعند إجراء المقارنات البعدية، اتفقت هذه المقارنات مع النتيجة السابقة.

ولقد أشارت النتائج بالإضافة لذلك عن وجود الأثر المعنوي لدلالة الفروق في تمييز الفقرة على اختبار الاختيار من متعدد باختلاف عدد البدائل وفقا للنظرية الحديثة في القياس، وذلك لصالح الاختبار ذي الخمس بدائل. وهذه النتيجة لاتتفق مع نتائج دراسة فان (Fan,1998).

وبعد إجراء المقارنات البعدية، وجد أن الاختبار ذي الخمس بدائل هو الأفضل في تقدير تمييز الفقرة وهذا متفق مع الأدب النظري الذي يشير الى أن زيادة عدد البدائل يزيد من القدرة التمييزية للفقرة.

كما أن هذه النتائج قد تم الحصول عليها في النظرية الكلاسيكية في القياس، فلقد أشارت النتائج الى وجود الأثر المعنوي لدلالة الفروق في تمييز الفقرة على اختبار الاختيار من متعدد باختلاف عدد البدائل، وذلك لصالح الاختبار ذي الخمس بدائل. وهذه النتائج تتفق مع دراسة ستراتون وكاتس (Straton & Catts,1980).

بالإضافة الى ذلك، أشارت النتائج الى وجود فروق ذات دلالة إحصائية في تخمين الفقرة على اختبار الاختيار من متعدد باختلاف عدد البدائل، وذلك لصالح الاختبار ذي الثلاث بدائل بسبب أن متوسط الاختبار

ذي الثلاث بدائل أكثر من متوسطات الاختبارات ذات الأربع والخمس بدائل. وهذه النتيجة أيضا تتفق مع كل الأدب النظري الذي يشير الى أن نسبة التخمين تنقص عند استخدام عدد أقل من البدائل.

التوصيات:

وتوصي هذه الدراسة باستخدام فقرات اختبار الاختيار من متعدد ذو الخمسة بدائل. ولقلة الدراسات الباحثة في هذا الموضوع، ولإعتماد معظم الدراسات على الطرق الإحصائية الوصفية، وعدم توظيف البيانات لفحص فرضيات احصائية بإستخدام الطرق الإحصائية التحليلية، يبقى هناك مجالا كبيرا للبحث في هذا المجال.

قائمة المراجع

أ- المراجع العربية

جلال، سعد. (1985). القياس النفسي المقاييس و الاختبارت. مكتبة المعارف الحديثة، الإسكندرية.

الدرابيع، ماهر. (2003). فعالية النموذج اللوغاريتمي ذي المعلمة الواحدة نموذج راش في دقة تقدير قدرة الفرد و معامل صعوبة الفقرة باختلاف حجم العينة و طول الاختبار. مجلة دراسات، الجامعة الأردنية، (قبلت للنشر).

كاظم، أمينة محمد. (1988). استخدام نموذج راش في بناء اختبار تحصيلي في علم النفس وتحقيق التفسير الموضوعي للنتائج. مؤسسة الكويت للتقدم العلمي، الكويت، ص (7-123).

علام، صلاح. (1995). الاختبارات التشخيصية مرجعية المحك. دار الأمل: القاهرة.

عودة، أحمد سليمان. (1993). القياس و التقويم في العملية التدريسية. ط2 ، دار الأمل، إربد.

عودة، أحمد سليمان و ملكاوي، فتحي حسن. (1992). أساسيات البحث العلمي، ط2، مكتبة الكتاني، إربد.

نشواتي، عبد المجيد. (1991). علم النفس التربوي، ط5، دار الفرقان، اربد.

ب- المراجع الأجنبية

Allen, M. J., & Yen, W. M. (1979). Introduction to Measurement Theory. California Cole Publishing Company.

Beker, F. B. (1985). The basics of item response theory. Portsmouth, NH: Heinemann.

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick, Statistical theories of mental test scores pp. 395-479. Reading, Ma: Addison-Wesley.

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. Psychometrika, vol. 46, pp. 443-459.

Bock, R. D., & Lieberman, M. (1970). Fitting a response model for an dichotomously scored items. Psychometrika, vol. 35, pp. 197-197.

Cambell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by multitrait-multimethod matrix. Psychological Bulletin, vol. 56, pp. 81-105.

Crocker, I., & Algina, j. (1986). Introduction to classical and modern test theory. New York: Holt, Rinehart & Winston.

Fan, X. (1998). Item response theory and classical test theory: An empirical comparison of their item/Person statistics. Educational and Psychological measurement, vol. 58, pp. 357-381.

Hambelton, R. K., & Cook, L. (1980). The robustness of the latent trait models and effects of test length and sample size on the precision of ability estimate. In D. J. Weiss (Ed), Proceedings of the 1979 computerized adaptive testing conference (pp. 36-52). Minneapolis, Minnesota.

Hambelton, R. K., & Traub, R. (1971). Information curves and efficiency of three logistic test models. British journal of mathematical and statistical psychology.

Hambelton, R. K., & Traub, R. (1973). Analysis of empirical data using two logistic latent trait models. British Journal of Mathematical and Statistical Psychology, vol. 26, pp. 195-211.

Hambelton, R. K., & Swamithan, H. (1985). Item Response Theory: Principles and Applications. Boston: Kluwer, Nijhoff Publishing.

Hambelton, R. K., Swamithan, H., & Rogers, H. J. (1991). Fundamentals of Item Response Theroy. Thousands Oaks, CA: Sage Publications.

Hambelton, R. K. & Jones, R. W. (1993). Comparison of classical test theory and response item theory and their applications to test devbelopment. Educational Measurment: Issues and Practice, vol. 12, pp. 38-47.

Harris, D. (1989). Comparison of 1-, 2-, and 3-parameter IRT models. Educational Measurement: Issues and Practice, vol. 8, pp. 35-41.

Harwell, M.R. (1997). Analyzing the results of Monte Carlo studies in item response theory. Educational and Psychological Measurement, vol. 57, pp. 266-279

Hulin, G. L., Lissak, R. I. & Drasgow, F. (1982). Recovery of two- and three- parameter logistic item characteristic curves: A monte carlo study. Applied Psychological Measurements, vol. 6, pp. 249-260.

Lord, F. and Novick, M. (1968). Statistical theories of mental test scores. Reading, Massachusitts: Addison-Wesley.

Lord, F. (1980). Applications of item response theory to practical testing problems. Hillsdale, NJ: Lawrence Erlbaum Associates.

Lord, F. (1952). A theory of test scores. Psychometric Monograph (No. 7).

Mislevy, R. J., & Bock, R. D. (1984). BILOG: Item analysis and test scoring with binary logistic models. Mooresville, IN: Scientific Software.

Owen, V. and Froman, D. (1987). What's wrong with three-option multiple choice items. Educational and Psychological Measurement, vol. 45,pp. 785-794.

Stevens, S. (1951). Mathmatics, Measurment, and psychophysiscs. In Stevens, S. (Ed.). Handbook of experimental Psychology New York: Wiley, 1951.

Straton, G. and Catts, M. (1980). A comparison of two, three, and four-choice item tests given a fixed total numbers of choices. Educational and Psychological Measurement, vol. 40, pp. 357-364.

Swaminathan, H., & Gifford, J. A. (1982). Bayesian estimation in the Rasch model. Journal of Educational Statistics, vol. 7, pp. 175-191.

Swaminathan, H., and Gifford, J. A. (1985). Bayesian estimation in the two-parameter logistic model. Psychometrika,vol. 50, pp. 349-364.

Warm, T. A. (1978). A primer of item response theory. Technical Report No. 940279. Oklahoma City: U.S. Coast Guard Institute.

Wingersky, M. S., Barton, M. A., & Lord, F. M. (1982). LOGIST user's guide. Princeton, NJ: Educational Testing Sevice.

Wright, B. D. (1977). Solving measurement problem with Rash Model. Journal of educational measurement, vol. 14, pp. 97-116.

Wright, B. D., & Stone, M. H. (1979). Best test design. Chicago: MESA Press.